



PCT/EP200 4 / 004 001

25. 05. 2004



INVESTOR IN PEOPLE

BEST AVAILABLE COPY

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

REC'D 22 JUN 2004

WIPO

PC

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

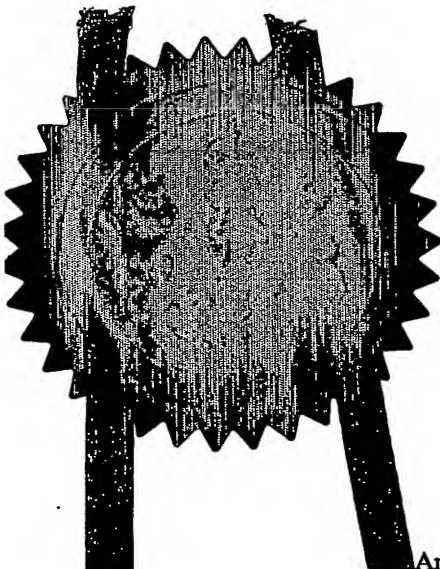
Signed

P. Mahoney

Dated 15 April 2004

**PRIORITY
DOCUMENT**

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)





16APR03 E800496-1 D02029
P01/7700 0.00-0308696.4

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office
Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference

GMPL/PMS/PB60213P

2. Patent application number

(The Patent Office will fill in his part)

15 APR 2003

0308696.4

3. Full name, address and postcode of the or of each applicant (underline all surnames)

Patents ADP number (if you know it)

If the applicant is a corporate body, give the country/state of its incorporation

Glaxo Group Limited
Glaxo Wellcome House, Berkeley Avenue,
Greenford, Middlesex UB6 0NN, Great Britain

00473 587003

United Kingdom
see continuation sheet for further applicant(s)

4. Title of the invention

New Process

5. Name of your agent (if you have one)

"Address for service" in the United Kingdom to which all correspondence should be sent
(including the postcode)

Patents ADP number (if you know it)

Corporate Intellectual Property

GlaxoSmithKline
Corporate Intellectual Property (CN9 25.1)
980 Great West Road
BRENTFORD
Middlesex TW8 9GS

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or each of these earlier applications and (if you know it) the or each application number

Country	Priority application number (if you know it)	Date of filing (day / month / year)
---------	-------------------------------------------------	----------------------------------------

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing (day / month / year)
-------------------------------	----------------------------------------

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? (Answer yes if:

- a) any applicant named in part 3 is not an inventor, or
 - b) there is an inventor who is named as an applicant, or
 - c) any named applicant is a corporate body
- See note (d)

- 9 Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	1	CF
Description	34	
Claim(s)	2	
Abstract	1	
Drawings	10 only	

10. If you are also filing any of the following, state how many against each item.

Priority Documents

Translations of priority documents

Statement of inventorship and right to grant of a patent (Patents Form 7/77)

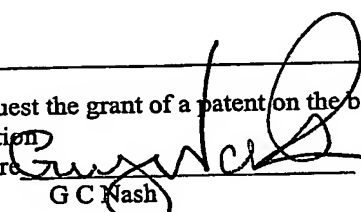
Request for preliminary examination and search (Patents Form 9/77)

Request for substantive examination (Patents Form 10/77)

Any other documents (please specify)

11.

We request the grant of a patent on the basis of this application

Signature  Date 15-Apr-03
G C Nash

12. Name and daytime telephone number of person to contact in the United Kingdom

G C Nash 01279 644284

Warning

After an application for a Patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission unless an application has been filed at least six weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

- If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505
- Write your answers in capital letters using black ink or you may type them.
- If there is not enough space for all relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.
- If you have answered 'Yes' Patents Form 7/77 will need to be filed.
- For details of the fee and ways to pay please contact the Patent Office.

CONTINUATION SHEET

Reference: GMPL/PMS/PB60213P

Further Applicant (s)

The Governors of the University of Alberta

c/o Industry Liaison Office, 222 Campus Tower, 8625-112 Street, Edmonton, Alberta

CANADA T6G 2J9

Canadian

05697719002

New Process

The present invention relates to new processes for improving the manufacture of clavams e.g. clavulanic acid. The present invention also provides novel DNA sequences and new microorganisms capable of producing increased amounts of clavulanic acid.

Microorganisms, in particular *Streptomyces sp.* produce a number of antibiotics including clavulanic acid and other clavams, cephalosporins, polyketides, cephamycins, tunicamycin, holomycin and penicillins. There is considerable interest in being able to manipulate the absolute and relative amounts of these antibiotics produced by the microorganism and accordingly there have been a large number of studies investigating the metabolic and genetic mechanisms of the biosynthetic pathways (Demain, A.L. (1990) "Biosynthesis and regulation of β -lactam antibiotics." in "50 years of Penicillin applications, history and trends").

Streptomyces clavuligerus produces two major groups of antibiotics; one being the cephamycins, cephalosporins and penicillins (Demain, A.L. (1990) *supra*) and the other comprising clavams. Clavams can be arbitrarily divided into two groups, 5S and 5R clavams, dependent on their ring stereochemistry. The commercially important clavam clavulanic acid, a component of the antibiotic Augmentin (trade mark of GlaxoSmithKline), is a 5R clavam. Examples of 5S clavams are clavam-2-carboxylate (C-2-C), 2-hydroxymethyl clavam (2HMC) and alanylclavam (Brown *et al.* (1979) J. Chem. Soc. Chem. pp282-283).

Genes encoding biosynthetic enzymes and regulatory proteins for clavulanic acid production have been located in a cluster next to the genes involved in cephamycin C production and make up a supercluster of antibiotic related genes within the *S. clavuligerus* genome (Alexander *et al.* (1998) J.Bacteriol. 180:4068-79). For example the genes encoding the enzymes involved in clavaminic acid production, a clavulanic acid precursor, which include *orf2* (*ceaS*) (Khaleeli *et al.* (1999) J. Am. Chem. Soc. 121:9223-9224), *orf3* (*bls*) (Bachmann and Townsend (1998) Chem. Commun.:2325-2326), *orf4* (*pah*) (Wu *et al.* (1995) J. Bacteriol. 177:3714-3720), *orf5* (*cas2*) (Marsh *et al.* (1992) Biochemistry. 31:12648-57) and perhaps *orf6* (Kershaw *et al.* (2002) Eur. J. Biochem. 269,2052-2059) are all located within the clavulanic acid cluster. Disruptions in *orfs2-6* cause a complete loss of clavulanic acid production when mutant cultures are grown on starch asparagine medium (Aidoo, K.A. *et al.* (1993) p219-236 In. V.P. Gullo, J.C. Hunter-Cevera, R. Cooper and R. K. Johnson (ed.), Developments in Industrial Microbiology series, vol.33 Society for Industrial Microbiology, Fredericksburg, Va.). However this loss is conditional upon the growth media used for when mutants are grown on Soy medium (Salowe *et al.* (1990) Biochemistry 29: 6499-6508) clavulanic acid production is partially restored (Jensen *et al.* (2002) Antimicrob. Agents and Chemother. 44: 720-726). This phenomenon could suggest that other genes present in the *S.*

clavuligerus genome could compensate in some way for the loss of the activity of these genes under certain conditions. Alternatively it could be that the Soy media contains very small amounts one or more of the metabolites produced by the *orfs* 2-6 allowing strains disrupted in these genes to make small amounts of clavulanic acid.

5 Marsh *et al.* (1992) *supra* has reported that *S. clavuligerus* contains two copies of the *cas* gene (*cas1* and *cas2*). *cas1* is not associated with the clavulanic acid gene cluster and has a high homology to *cas2*. Disruption of *cas2* decreases clavulanic acid production by 35% when cultures are grown on Soy medium and eliminates production entirely when cultures are grown on starch asparagine (SA) medium (Paradkar and Jensen 1995 J.Bact 177: 1307-10 1314). The disruption of the *cas1* gene results in mutants which produce near wild type levels of clavulanic acid on SA medium, but produce 31-73% less clavulanic acid when grown on Soy medium than the wild type (Mosher *et al* (1999) Antimicrob. Agents and Chemother. 43: 1215-1224). It is also reported that in mutant strains where both the *cas1* and *cas2* genes have been disrupted no clavulanic acid is produced under any of the fermentation conditions tested. 15 Interestingly when the genes surrounding *cas1* were sequenced, no additional genes involved in clavulanic acid production were found but instead six novel genes involved in 5S clavam biosynthesis (named *cvm1* to 6) were identified. (Mosher *et al* (1999) *supra*). Further work on these 5S clavam-specific genes showed that disruption of the genes, using genetic engineering methodologies, leads to improvements in the levels of clavulanic acid made by 20 the mutant strains and also dramatic reductions in the levels of 5S clavam production (WO98/33896). This reduction in 5S clavam production, in particular the 5S clavam clavam-2-carboxylate, is especially important in the commercial production of clavulanic acid because some 5S clavams are known to be toxic and for this reason the levels are tightly controlled within the British and US Pharmacopoeias.

25 Despite these advances in the understanding of clavulanic acid biosynthesis it is still a highly desirable goal in the pharmaceutical industry to continue to improve production methods for clavulanic acid, both for reasons of cost and for reasons of safety.

The following definitions are provided to facilitate understanding of certain terms used frequently herein:

30 "Gene" as used herein also includes any regulatory region required for gene function or expression.

"*cvm*" genes as used herein refers to any of the genes *cvm1*, *cvm2*, *cvm3*, *cvm4*, *cvm5*, *cvm6* or *cvm7* as defined hereinabove.

"*cvmpara*" genes as used herein refers to any of the genes *cvm6para* or *cvm7para* as 35 defined hereinabove.

"*orf*" genes as used herein refers to any of the genes *orf2*, *orf3*, *orf4*, *orf5*, *orf6*, *orf7*, *orf8*, *orf9*, *orf10*, *orf11*, *orf12*, *orf13*, *orf14*, *orf15*, *orf16*, *orf17*, or *orf18* as defined hereinabove.

"*orfpara*" genes as used herein refers to any of the genes *orf2para*, *orf3para*, *orf4para* or *orf6para* as defined hereinabove.

"Disrupted" as used herein means that the activity of the gene (with regard 5S clavam production) has been reduced or eliminated by, for example, insertional inactivation using an antibiotic resistance gene, preferably apromycin (Paradkar, A.S and Jensen, S.E (1995) *supra*), or other mutagenesis technique (for example those disclosed in Sambrook et al (1989) *supra*). Other mutagenesis techniques include insertion of other DNAs (not antibiotic resistance genes), site-directed mutagenesis to either change one or more bases in the gene sequence or insert one or more bases into the sequence of the gene.

"Deleted" as used herein means that the gene, or a segment thereof, has been deleted (removed) from a larger polynucleotide which, before the deletion was performed, included said gene or segment thereof. When the polynucleotide bearing the deletion is introduced into the genome of the microorganism by means of gene replacement technology (Paradkar and Jensen (1995) *supra*) the activity of the gene or protein encoded thereby is eliminated or reduced such that the levels of 5S clavam produced by the microorganism are reduced. The deletion may be large (for example the complete open reading frame with or without regulatory control regions) or small (for example a single base pair resulting in a frameshift mutation).

"Reduced" as used herein means that the levels of 5S clavam produced by the microorganism of the invention are lower than the levels produced in the corresponding *S. clavuligerus* strain which has not had the relevant open reading frames disrupted or deleted. The corresponding *S. clavuligerus* is therefore the "parent" strain into which the disrupted or deleted open reading frames were subsequently introduced to generate the microorganism of the invention.

"At least maintained" as used herein means that the level of clavulanic acid produced in the microorganism of the invention is the same or greater than that produced in the corresponding *S. clavuligerus* strain which has not had the relevant open reading frames disrupted or deleted. The corresponding *S. clavuligerus* is therefore the "parent" strain into which the disrupted or deleted open reading frames were subsequently introduced to generate the microorganism of the invention.

The present invention concerns new processes for making clavulanic acid using newly identified *S. clavuligerus* genes. Using a probe derived from *orf4*, a fragment of the *S. clavuligerus* genome has been isolated and has been shown to comprise a number of genes that when disrupted are shown to affect 5S and 5R clavam biosynthesis in *S. clavuligerus*. Sequence analysis of the fragment has indicated the presence of a gene showing high similarity to *orf4* (hereinafter called *orf4par*). However surprisingly further sequence analysis of the regions flanking the *orf4par* gene has revealed a new cluster of genes

comprising paralogues of genes previously identified in both the clavulanic acid (*cas2* cluster) and 5S clavam (*cas1* cluster) gene clusters.

Accordingly the invention provides a *S. clavuligerus* microorganism comprising DNA corresponding to one or more open reading frames essential for 5S clavam biosynthesis, wherein said open reading frames are disrupted or deleted such that the production of 5S clavams by said *S. clavuligerus* is reduced and clavulanic acid production is at least maintained, wherein the open reading frames are selected from:

- a) *cvm6para* (SEQ ID NO:1);
- b) *cvm7para* (SEQ ID NO:2);
- c) *cvm6para* and *cvm6* (SEQ ID NO:5); or
- d) *cvm7para* and *cvm7* (SEQ ID NO:6).

In a second aspect the invention provides a *S. clavuligerus* microorganism comprising DNA corresponding to one or more open reading frames essential for 5S clavam biosynthesis, wherein said open reading frames are disrupted or deleted such that the production of 5S clavams by said *S. clavuligerus* is reduced and clavulanic acid production is at least maintained, wherein the open reading frames are selected from:

- a) *cvm6para* and one or more of *cvm1* (SEQ ID NO:7), *cvm2* (SEQ ID NO:8), *cvm3* (SEQ ID NO:9), *cvm4* (SEQ ID NO:10), *cvm5* (SEQ ID NO:11), *cvm6*, *cvm7* or *cvm7para*; or
- b) *cvm7para* and one or more of *cvm1*, *cvm2*, *cvm3*, *cvm4*, *cvm5*, *cvm6*, *cvm7* or *cvm6para*.

The genes *cvm1*, *cvm2*, *cvm3*, *cvm4*, *cvm5* and *cvm6* are disclosed in Mosher et al (1999) *supra* and WO98/33896 (*cvm1* is *orfup1*, *cvm2* is *orfup2*, *cvm3* is *orfup3*, *cvm4* is *ordwn1*, *cvm5* is *orfdwn2* and *cvm6* is *orfdwn3*). The *cvm7* gene, found to be a further 5S clavam specific gene of the 5S clavam (*cas1*) cluster, has been identified during work leading to the present invention and is disclosed hereinbelow.

In a further aspect the invention provides isolated polynucleotides comprising the *cvm6para* and *cvm7para* open reading frames which are used in the preparation of the *S. clavuligerus* microorganism of the invention. Preferably said polynucleotides comprise open reading frames selected from the group consisting of:

- a) *cvm6para*;
- b) *cvm7para*;
- c) *cvm6para* and *cvm6*;
- d) *cvm7para* and *cvm7*;
- e) *cvm6para* and one or more of *cvm1*, *cvm2*, *cvm3*, *cvm4*, *cvm5*, *cvm6*, *cvm7* or *cvm7para*; or
- f) *cvm7para* and one or more of *cvm1*, *cvm2*, *cvm3*, *cvm4*, *cvm5*, *cvm6*, *cvm7* or *cvm6para*.

In another aspect the present invention provides vectors for cloning and manipulating the *cvm* polynucleotides disclosed herein and which can be used in the preparation of the *S.*

clavuligerus microorganism of the invention. Processes for using these vectors to make the *S. clavuligerus* microorganism of the invention are also provided.

The encoded polypeptides from *cvm6para* and *cvm7para* are also provided by the invention (SEQ ID NO:3 and SEQ ID NO:4 respectively).

5 The invention further provides a polynucleotide comprising one or more open reading frames encoding one or more enzymes involved in clavulanic acid biosynthesis wherein said open reading frames are selected from the group consisting of:

- a) *orf2para* (SEQ ID NO:12),
- b) *orf3para* (SEQ ID NO:13),
- 10 c) *orf4para* (SEQ ID NO:14), and
- d) *orf6para* (SEQ ID NO:15).

In a further aspect the invention provides a polynucleotide comprising one or more open reading frames encoding one or more enzymes involved in clavulanic acid biosynthesis wherein said open reading frames comprise one or more of:

- 15 a) *orf2para*,
- b) *orf3para*,
- c) *orf4para*,
- d) *orf6para*

in combination with one or more genes involved in clavulanic acid biosynthesis selected from
 20 *orf2*, *orf3*, *orf4*, *orf5*, *orf6*, *orf7*, *orf8*, *orf9*, *orf10* (Canadian patent application CA2108113 and Jensen, S.E et al (2000) Antimicrob. Agents Chemother 44:720-6) *orf11*, *orf12* (Li, R.N et al (2000) J. Bacteriol 182:4087-95), *orf13*, *orf14*, *orf15*, *orf16*, *orf17*, or *orf18* (patent application PCT/GB02/04989).

Vectors comprising such polynucleotides are also provided by the present invention
 25 together with processes for the use of such vectors to prepare strains of *Streptomyces clavuligerus* which can be used to produce elevated levels of clavulanic acid.

Strains of *Streptomyces clavuligerus* so produced and methods for using them to produce clavulanic acid by fermentation are also provided.

Thus the invention further provides a *Streptomyces clavuligerus* microorganism
 30 comprising a vector comprising a polynucleotide comprising one or more open reading frames encoding one or more enzymes involved in clavulanic acid biosynthesis wherein said open reading frames are selected from the group consisting of:

- a) *orf2para*,
- b) *orf3para*,
- 35 c) *orf4para*, and
- d) *orf6para*.

In a further aspect the invention provides a *Streptomyces clavuligerus* microorganism comprising a vector comprising a polynucleotide comprising one or more open reading frames encoding one or more enzymes involved in clavulanic acid biosynthesis wherein said open reading frames are selected from the group consisting of:

- a) *orf2para*,
- b) *orf3para*,
- c) *orf4para*,
- d) *orf6para*

in combination with one or more genes involved in clavulanic acid biosynthesis selected from *orf2*, *orf3*, *orf4*, *orf5*, *orf6*, *orf7*, *orf8*, *orf9*, *orf10* (Canadian patent application CA2108113 and Jensen, S.E et al (2000) Antimicrob. Agents Chemother 44:720-6) *orf11*, *orf12* (Li, R.N et al (2000) J. Bacteriol 182:4087-95), *orf13*, *orf14*, *orf15*, *orf16*, *orf17*, or *orf18* (patent application PCT/GB02/04989).

The present invention also contemplates a *S. clavuligerus* micororganism comprising a combination of one or more disrupted or deleted *cvm6para* or *cvm7para* genes, optionally in combination with other disrupted or deleted 5S genes previously disclosed, together with vectors comprising *orf2para*, *orf3para*, *orf4para* or *orf6para* genes, optionally in combination with other clavulanic acid biosynthetic genes (selected from the genes *orf2* to *orf18*) previously disclosed.

Polynucleotides of the invention can be isolated by conventional cloning methods, such as PCR or library screening methods, using the sequences disclosed herein and in Mosher et al (1999) *supra*, WO98/33896, Canadian patent application CA2108113, Jensen, S.E et al (2000) *supra*, Li, R.N et al (2000) *supra* and patent application PCT/GB02/04989, as indicated hereinabove. Examples of such cloning methods are described in, for example, Sambrook, J et al (1989) Molecular cloning, a laboratory manual (2nd Ed) Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Polynucleotides comprising individual open reading frames can be isolated and ligated together into vectors in a variety of combinations as defined hereinabove using techniques well know in the art. The choice of vector will depend on the function being carried out, for example cloning, expression, gene inactivation or transfer into *S. clavuligerus* eg. for gene replacement. In all cases a variety of vectors are available to the skilled person and are well known in the art. For example such vectors are known from Sambrook, J et al (1989) *supra* for general cloning vectors Hopwood, D.A et al (1985) *supra* for Streptomyces vectors, Paradkar and Jensen (1995) *supra*, Mosher et al (1999) *supra* and WO98/33896 *supra* for gene disruption and gene replacement vectors and CA2108113 *supra* for vectors suitable for expression of genes in *Streptomyces clavuligerus*. However the choice of vector is not limited to just those disclosed in these sources.

Further, in the case of the gene combinations involving the *orf2para*, *orf3para*, *orf4para*, *orf5para* and *orf6para* genes the skilled artisan would be able to design suitable DNA constructs

to ensure that each open reading frame is suitably positioned relative to a transcriptional promoter, whether this be the native promoter or a heterologous promoter that also functions in the *Streptomyces clavuligerus* background, or indeed other regulatory sequence, in such a manner that expression of each open reading frame is optimally achieved.

Subsequent manipulation of the polynucleotides, in particular with respect their introduction into the *Streptomyces clavuligerus* background, can be carried out according to standard methods as disclosed in, for example, Hopwood, D.A et al (1985) *supra*. Disruption of gene sequences, and subsequent gene replacement, can be carried out according to the method of Paradkar, A.S and Jensen, S.E (1995) *supra*. Deletion of gene sequences can be carried out using well established techniques, for example that disclosed in WO98/33896.

Microorganisms of the invention can be prepared from *Streptomyces clavuligerus* strains including, but not limited to, *Streptomyces clavuligerus* ATCC 27064 (American Type Culture Collection, Manassas, Virginia, USA), alternatively available as NRRL 3585 (Northern Regional Research Laboratory, Peoria, Illinois, USA). For example mutant strains of *Streptomyces clavuligerus* can also be used including those prepared by genetic engineering techniques, or those prepared by strain improvement methods. Examples of such strains include *Streptomyces clavuligerus* strains 56-1A, 56-3A, 57-2B, 57-1C, 60-1A, 60-2A, 60-3A, 61-1A, 61-2A, 61-3A or 61-4A as disclosed in WO98/33896.

Thus in another aspect the invention relates to a process for improving clavulanic acid production in a suitable microorganism comprising isolating a polynucleotide as described hereinabove, manipulating said polynucleotide, introducing the manipulated polynucleotide into a said suitable microorganism and fermenting said suitable microorganism under conditions whereby clavulanic acid is produced. Manipulation of said polynucleotide may be by means of disrupting or deleting gene sequences in the case of *cvm* genes, optionally together with *cvm* genes, or by inserting into vectors suitable for expression in the case of *orf* genes, optionally together with *orf* genes.

Preferably the suitable microorganism is *Streptomyces clavuligerus*.

Such fermentation methods are well known in the art, for example the methods disclosed in UK Patent Specification No. 1,508,977. Methods for using clavulanic acid in the preparation of antibiotic formulations are similarly well known in the art.

Examples

Example 1 - Materials and Methods

In the examples all methods are as described in Sambrook, J. et al *supra*, Hopwood, D.A. et al. (1985) *supra* and Kieser, T et al. (2000) *Practical Streptomyces Genetics*, unless otherwise stated. Transformation methods can also be found in Paradkar, A.S. and Jensen, S.E (1995) *supra*.

1.1 Bacterial strains, media and culture conditions.

Streptomyces clavuligerus NRRL 3585 was obtained from the Northern Regional Research Laboratory (Peoria, IL). *S. clavuligerus* was maintained on either MYM agar (Stuttard, C. (1982) J. Gen. Microbiol. 128:115-121) or ISP Medium #4 agar plates (Difco, Detroit, MI).

Cultures for the isolation of chromosomal DNA were grown on a 2:3 mixture of trypticase soy broth and YEME as described by Alexander et al. (1998) J. Bact. 180:4068-79. Cultures for analysis of the production of clavulanic acid and other clavam metabolites were grown on Soy medium (European Patent 0349 121) unless otherwise stated. All liquid cultures were grown at 26°C on a rotary shaker at 250 rpm.

Manipulation of DNA in *Escherichia coli* was done using strain XL-1 Blue (Stratagene, La Jolla, CA). *E. coli* cultures were maintained on LB agar medium and grown in liquid culture in LB medium at 37°C (Sambrook, J et al (1989) *supra*). Plasmid-containing cultures were supplemented with appropriate levels of antibiotic.

1.2 DNA manipulations.

Standard DNA manipulations such as plasmid isolation, restriction endonuclease digestion, generation of blunt-ended fragments, ligation, ³²P labelling of DNA probes by nick translation and *E. coli* transformation were carried out as described in Sambrook J et al (1989) *supra*. Plasmid and genomic DNA isolation from *Streptomyces* spp. was conducted as described in Kieser, T et al (2000) *supra*. Construction of a library of *S. clavuligerus* genomic DNA fragments in the cosmid pWE15 was carried out according to the manufacturer's instructions (Stratagene).

Southern analysis of *S. clavuligerus* DNA fragments was conducted at high stringency as described by Sambrook, J et al (1989) *supra*. Hybridization membranes were washed twice for 30 min at 2xSSC/0.1% SDS and once for 30 min at 0.1xSSC/0.1% SDS, all at 65°C.

Example 2 - Preparation of the paralogue cluster DNA fragment

2.1 Cloning and nucleotide sequencing of the *orf4* paralogue

A strong and a very weak hybridization signal was consistently observed on Southern blots of *Nco*I-digested *S. clavuligerus* chromosomal DNA when probed with the *orf4* gene (CA2108113). The strong signal corresponded to the *orf4* gene, but the identity of the gene that gave rise to the very weak signal was unknown. Therefore it was decided to clone this gene. To this end, *Nco*I fragments from *S. clavuligerus* DNA of approximately 4-5kb in size were ligated into *Nco*I digested pUC120 (Vieira, J and J Messing (1987) Methods Enzymol.

153, 3-11) and screened using a colony blot hybridisation method and employing the *orf4* gene as a probe. Plasmid DNA was isolated from potential positive clones and confirmed to carry a 4.3 kb *NcoI* fragment. A representative clone, pO4H-4, was chosen for further study. The sequencing of the 4.3 kb *NcoI* fragment was carried out. Analysis of the sequence generated identified three genes, one which had homology to *orf4* and was called *orf4par*. The two other genes present were found to have homology with *orf6* and *cvm6* and were therefore called *orf6par* and *cvm6par*. This result suggested that this region of DNA may contain a cluster of genes with paralogues in either the clavulanic acid biosynthetic gene cluster or the *cvm* clavam biosynthetic gene cluster.

2.2 Sequencing of DNA flanking the 4.3 kb *NcoI* fragment containing *orf4par*

Sequence analysis of DNA flanking the 4.3 kb *NcoI* fragment containing *orf4par* was achieved by identifying 2 cosmid clones containing the *orf4par* gene. The two cosmid clones containing *orf4par*, 14E10 and 6G9, were isolated from a *S. clavuligerus* pWE15 (Promega, Madison, WI) cosmid bank that had been probed with a 0.46Kb *SaII* fragment that is internal to the *orf4par* gene. These cosmids have been partially mapped using a series of digestions and Southern hybridization experiments (In. Nucleic acid techniques in bacterial systematics. Ed. Stackebrandt, E and Goodfellow, M (1991) John Wiley and Sons, p205-248). Digestion of both cosmids with *EcoRI*, *KpnI* and *NruI* suggest that the insert size of 14E10 is approximately 45 kb and 6G9 is approximately 40 kb. These two cosmid inserts have about 20 kb of overlapping DNA and provided DNA for sequence analysis of regions upstream and downstream of the 4.3 kb *NcoI* fragment containing *orf4par*.

DNA sequence information was generated essentially as described in CA2108113. The DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Pharmacia, Baie d'Urfe, Quebec, Canada) was used. Approximately 13.3 kilobases of contiguous DNA sequence was generated. The nucleotide sequence of the *S. clavuligerus* chromosomal DNA generated in these experiments is shown in SEQ ID No:16.

A number of open reading frames were identified which displayed significant homology with the previously described *orf2*, *orf3*, *orf4*, and *orf6* (CA2108113). These genes have been located within the genome in relation to each other, and are found to be nearly in the same organisation as that of the genes within the clavulanic acid cluster. The genes *orf2par*, *orf3par* and *orf4par* are adjacent to each other and in the same orientation as their counterparts *orf2*, *orf3* and *orf4*. However *cas1* is not downstream of *orf4par* as *cas2* is to *orf4* in the clavulanic acid pathway but is instead within the clavam cluster (Mosher et al (1999) *supra*). Another difference between the clavulanic acid cluster and the paralogue arrangement is that *orf6par* is end-on-end to *orf4par*, and so is not in the same orientation as *orf2par-4par*, whereas *orf6* is in the same orientation as *orfs2-4* in the clavulanic acid cluster.

Suprisingly the gene immediately upstream of *orf6par*, was found to be a gene that had a paralogue in the clavam and not the clavulanic acid cluster. This gene was called *cvm6par*, as it is a paralogue of the *cvm6* gene found clustered with *cas1* (Mosher et al (1999) *supra*). The *cvm6* gene encodes an enzyme that is involved in clavam production (*orfdwn3* in WO98/33896).

Located adjacent to *cvm6par* is a new gene called *cvm7par*. This gene shows homology to *cvm7*, a gene that is located upstream of *cvm3* in the clavam cluster (further described hereinbelow). Upstream of *cvm7* is a new open reading frame, believed to encode a sensor kinase. It encodes a polypeptide of 555 amino acids and shows good similarity to sensor kinase domains of two component response regulator genes.

2.3 Functional analysis of the open reading frames

Computer analysis of the DNA sequence shown in SEQ ID No.16 predicts the presence of 7 open reading frames. A description of each gene is shown in Table 1.

Table 1

Orf Designation	Homology (blast P)
<i>orf2par</i>	acetolactate synthase (67% identity to <i>orf2</i> carboxyethyl arginine synthase CEAS)
<i>orf3par</i>	asparagine synthetase (49% identity with <i>orf3</i> β -lactam synthase BLS)
<i>orf4par</i>	amidinohydrolase (71% identity with <i>orf4</i> amidinohydrolase PAH)
<i>orf6par</i>	ornithine acetyltransferase (47% identity with <i>orf6</i> ornithine acetyl transferase OAT)
<i>cvm6par</i>	aminotransferase (66% identity with <i>cvm6</i> acetylornithine aminotransferase)
<i>cvm7par</i>	Transcriptional regulator (33% identity with <i>cvm7</i> homologue)

Sensor Kinase	Sensor Kinase 47% identity with 2 component system from <i>S.coelicolor</i> A3 (2)
---------------	------------------------------------------------------------------------------------------

To assess the possible roles of these ORFs in the biosynthesis of clavulanic acid and/or clavams produced by *S. clavuligerus*, insertional inactivation mutants were created by gene replacement essentially as described by Paradkar and Jensen (1995) *supra*.

- 5 However, in order to definitively define the phenotype of these disruptions, it was considered important to disrupt *orf3par*, *orf4par*, *orf6par* and *cvm6par* not only in wild type *S. clavuligerus*, but also in strains of *S. clavuligerus* that were already defective in the expression of *orf3*, *orf4*, *orf6*, and *cvm6* respectively. The *orf3,4* and 6 mutants were made as described in United States Patent No. 6,332,106 and the *cvm6* mutant made as described in
10 WO98/33896.

Example 3 – Analysis *orf4*, and *orf4par*

3.1 Construction of *orf4* mutants

- 15 Mutants disrupted in *orf4* (*pah*) were made as described in United States Patent No. 6,332,106.

3.2 Construction of *orf4par* mutants

- pO4H-4 (4.3kb *NcoI* fragment cloned into the *NcoI* site of pUC120 (Vieira and Messing 1987 *supra*) was digested with *KpnI* (one site in the cloned fragment and one site in the
20 vector) and religated to reduce the size of the *orf4par*-bearing DNA insert to 1.7kb thereby generating the plasmid p4K-1. The *orf4par* gene within p4K-1 was disrupted by digestion at its centrally located *EcoNI* site and insertion of the apramycin (*apr*) resistance gene cassette from pUC120apr (Trepanier et al. (2002) Microbiology 148: 643-656) after both fragments
25 had been made blunt by treatment with the Klenow fragment of DNA polymerase I. The *KpnI/NcoI* insert carrying the disrupted *orf4par* gene was then inserted into the *EcoRI* site of pDA501 after blunting the ends of both insert and vector. pDA501 is a shuttle vector prepared by fusing the *Streptomyces* plasmid pIJ486 (Kieser, T et al (2000) *supra*) to the *E.coli* plasmid pTZ18R (Stratagene) by means of their *EcoRI* and *BamHI* sites. The resulting construct, 6pDAB, was used to transform *S.lividans* TK24, and finally wild-type *S.*
30 *clavuligerus* to thiostrepton (thio at 5µg/ml) and apramycin (*apr* at 20µg/ml) resistance.
- Gene replacement mutants were generated as described by Paradkar and Jensen (1995) *supra*.

3.3 Construction of *orf4/orf4par* mutants

An approach was undertaken to generate the double mutant by transforming protoplasts of the *orf4par* (*apr^r*) mutant with the *orf4* (*thio^r*) disruption construct (Aidoo *et al.* (1994) Gene. 147:41-6). Protoplast preparations from *orf4par* mutants, were transformed with the *orf4* disruption construct isolated from *S.lividans*. Transformants were selected on thiostrepton at 5µg/ml and hygromycin (*hyg*) at 50µg/ml. Primary transformants were put through two rounds of sporulation under non- selective conditions in order to generate gene replacement mutants as described by Paradkar and Jensen (1995) *supra*.

3.4 Fermentation analysis of *orf4*, *orf4par* and *orf4/orf4par* mutants

To test the effect of disrupting *orf4*, *orf4par* and *orf4/orf4par* on clavulanic acid biosynthesis, spores from each isolate were inoculated into 20ml of seed medium (European patent 0 349 121) and grown for 2 days at 26°C with shaking. 1ml of the seed culture was then inoculated into a final stage Soy medium (European Patent 0349 121) and grown at 26°C for up to 3 days with shaking. Samples of final stage broth were withdrawn after three days growth and assayed for clavulanic acid productivity by HPLC (Mosher *et al* (1999) *supra*) and/ or using an imidazole derivatised colorimetric assay (Bird, A.E. *et al* (1982) Analyst, 107: 1241-1245 and Foulston, M. and Reading, C. (1982) Antimicrob. Agents Chemother., 22:753-762).

Fermentation analysis of *orf4* disruptant

The *orf4* disruptant was fermented in Soy medium and compared to wild type *S. clavuligerus* for production of clavulanic acid. After 72hrs growth, accumulation of clavulanic acid was reduced by 71%.

From these results it can be concluded that *orf4* is required for efficient production of clavulanic acid as elimination of this gene by disruption causes a reduction in clavulanic acid levels.

Fermentation analysis of *orf4par* disruptant

Mutant 5pDA defective in the *orf4par* gene was fermented in Soy medium and compared to wild type *S. clavuligerus* for production of clavulanic acid. After 72hrs growth, accumulation of clavulanic acid was reduced by 12%.

From these results it can be concluded that, like *orf4*, *orf4par* contributes to clavulanic acid biosynthesis as elimination of this gene by disruption causes a reduction in clavulanic acid levels.

Fermentation analysis of *orf4/orf4par* disruptants

When mutants A4-A1 and 3A3-A3, defective in both copies of the *orf4* genes were grown in Soy medium production of clavulanic acid could not be detected.

From these results it can be concluded that under the conditions tested, both genes, *orf4* and *orf4par*, contribute to clavulanic acid biosynthesis as the double disruption, results in a mutant unable to make clavulanic acid.

5 3.5 Southern Analysis

The *orf4*, *orf4par* and *orf4/4par* mutants were further characterised by Southern analysis. The results confirmed that in these mutants the chromosomal copies of the relevant genes had been disrupted as expected.

10 **Example 4 - Analysis of *orf6* and *orf6par***

4.1 Construction of *orf6* mutants

orf6 mutants were made as described in United States Patent No. 6,332,106

4.2 Construction of *orf6par* mutants

15 The *orf6par* gene was disrupted by introduction of a neomycin resistance gene (*neo^r*) into the *RsrII* site, approximately midway through the coding region. In order to achieve this pO4H-4 was digested with *KpnI* to remove *orf4par* and self ligated to give p5K-6. p5K-6 was digested with *RsrII* and the neomycin resistance gene, released from pFDNeo-S (Denis and Brzezinski (1992) Gene 111:115-118.) as a *PstI/EcoRI* fragment, was inserted after both
20 fragments had been made blunt by treatment with the Klenow fragment of DNA polymerase I. The construct pNeo5K-6A was obtained which has the *neo^R* gene in the same orientation as the *orf6par* gene.

A shuttle vector called pNeo5K-6Atsr#14 was constructed by inserting pIJ486 as a 6.2 Kb fragment linearised with *BglIII*, into the *BamHI* polylinker site of pNeo5K-6A. The
25 shuttle vector was used to transform *S. lividans* TK24 and finally *S. clavuligerus* WT to thiostrepton (5µg/ml) and neomycin (50µg/ml) resistance. Primary transformants were subjected to two rounds of sporulation under non- selective conditions in order to generate gene replacement mutants as described by Paradkar and Jensen (1995) *supra*.

30 4.3 Construction of *orf6/orf6par* mutants

orf6/orf6par double mutants were generated by transforming protoplasts of the *orf6par* (*neo^r*) mutant with the *orf6(apr^r)* disruption construct (Mosher et al (1999) *supra*). Protoplast preparations from *orf6par* mutants, were transformed with the *orf6* disruption construct isolated from *S.lividans*. Transformants were selected on apramycin (*apr*) at 50µg/ml. Primary
35 transformants were put through two rounds of sporulation under non- selective conditions in

order to generate gene replacement mutants as described by Paradkar and Jensen (1995) *supra*.

4.4 Fermentation of *orf6*, *orf6par* and *orf6/orf6par* mutants

To test the effect of disrupting *orf6*, *orf6par* and *orf6/orf6par* on clavulanic acid biosynthesis, spores from each isolate were tested as previously described in section 3.4.

Fermentation Analysis of *orf6* mutants

Mutant 6-1A defective in the *orf6* gene was fermented in Soy medium and compared to wild type *S. clavuligerus* for production of clavulanic acid. After 72hrs growth, accumulation of clavulanic acid was reduced by 57%. From these results it can be concluded that *orf6* is required for efficient production of clavulanic acid as elimination of this gene by disruption causes a reduction in clavulanic acid levels.

Fermentation Analysis of *orf6par* mutants

Mutant 14-2B(2) defective in the *orf6par* gene was fermented in Soy medium and compared to wild type *S. clavuligerus* for production of clavulanic acid. After 72hrs growth, accumulation of clavulanic acid was reduced by 27%. From these results it can be concluded that, like *orf6*, *orf6par* contributes to clavulanic acid biosynthesis as elimination of this gene by disruption causes a reduction in clavulanic acid levels.

Fermentation Analysis of *orf6/orf6par* mutants

Two separate mutants defective in both *orf6* and *orf6par* were fermented in Soy medium and compared to wild type *S. clavuligerus* for production of clavulanic acid. After 72hrs growth, accumulation of clavulanic acid was reduced by an average of 65%.

From these results it can be concluded that both *orf6* and *orf6par* are necessary for efficient production of clavulanic acid since disruption of either copy of the gene causes a reduction in clavulanic acid production. Inactivation of both copies of the gene caused a further decrease, but not a complete loss of clavulanic acid producing ability.

4.5 Southern Analysis

The *orf6*, *orf6par* and *orf6/orf6par* mutants were further characterised by Southern analysis. The results confirmed that in these mutants the chromosomal copy of the relevant gene had been disrupted as expected.

Example 5 – Analysis of *cvm6* and *cvm6par*

5.1 Construction of *cvm6* mutants

Construction of mutants disrupted in *cvm6* has already been described in WO98/33896 (*cvm6* is *orf4wn3*).

5.2 Construction of *cvm6par* mutants

A 1.7 Kb *SaII* fragment containing *cvm6par* was released from pO4H-4 and ligated into pUC118 at the *SaII* site. The resulting plasmid was digested with *EcoNI* to release a 140 bp fragment internal to *cvm6par*. In place of this fragment, the neomycin resistance gene from pFDNeo-S, released as an *EcoRI/PstI* fragment, was ligated into *cvm6par* after both fragments had been made blunt by treatment with the Klenow fragment of DNA polymerase I. The neo^R marker was inserted in the same orientation as *cvm6par*. The neomycin containing *SaII* fragment was released with *EcoRI* and inserted into the shuttle vector pUWL-KS (Weimeier, U.F (1995) Gene 165:149-150.) at the *EcoRI* site. The construct was named pNeoSal1.7U.

The plasmid pNeoSal1.7U was used to transform *S.lividans* TK24, and finally *S.clavuligerus* wild type. The resulting *cvm6par::neo* transformants were selected on MYM medium with 50µg/ml neomycin and 5µg/ml thiostrepton and then subjected to two rounds of sporulation under non- selective conditions to give double cross-over mutants.

5.3 Construction of *cvm6/cvm6par* mutants

The construct pNeoSal1.7U isolated from *S.lividans* TK24 was also used to transform the *cvm6* mutant 56-3A, where the apr^R cassette was inserted into *cvm6* in the same orientation as the gene. Transformants were grown on MYM medium with 50µg/ml neomycin and 5µg/ml thiostrepton. The mutants were put through two rounds of sporulation under non- selective conditions as described above and double cross-over mutants were isolated.

5.4 Fermentation of *cvm6*, *cvm6par* and *cvm6/cvm6par* mutants

To test the effect of disrupting *cvm6*, *cvm6par* and *cvm6/cvm6par* on β-lactam biosynthesis, spores from each isolate were tested as previously described in section 3.4.

Fermentation Analysis of *cvm6* mutants

It was reported in WO98/33896 that mutants 56-1A, 56-3A, 57-1C and 57-2B defective in the *cvm6* gene produced elevated levels of clavulanic acid (125-141% of the control strain) and greatly reduced levels of clavam-2-carboxylate and 2-hydroxymethylclavam when cultured in Soy medium.

These results suggest that the *cvm6* gene is required for efficient production of the 5S clavams. Disruption of *cvm6* not only results in a reduction in clavams but also a simultaneous increase in clavulanic acid.

Fermentation Analysis of *cvm6par* mutants

Mutants 3A1, 3A2, 2A-6, 2B-1 and 2B-2 defective in the *cvm6par* gene were fermented in Soy medium and compared to wild type *S. clavuligerus* for production of β-lactam

metabolites. After 72hrs growth, accumulations in clavulanic acid were increased by 6-11%. Production of clavam-2-carboxylate and alanyl clavam was abolished and levels of 2-hydroxymethyl clavam reduced by 50-85%.

These results suggest that like *cvm6* the *cvm6par* gene is required for efficient production of the 5S clavams. Disruption of *cvm6par* not only results in a reduction in clavams but also a simultaneous increase in clavulanic acid.

Fermentation Analysis of *cvm6/cvm6par* double mutants

Mutants A-1, A-2, B-1, B-2, C-1 and C-2 defective in both the *cvm6* and *cvm6par* genes were grown in Soy medium and compared to wild type *S. clavuligerus* for their production of β -lactam metabolites. Production of clavulanic acid was increased by 12-27%, production of alanyl clavam and clavam-2-carboxylate eliminated and levels of 2-hydroxymethyl clavam reduced by 70-83%.

These results indicate that, like the *cvm6* and *cvm6par* single mutants, the *cvm6/cvm6par* double mutants produced elevated levels of clavulanic acid and both genes are required for the efficient production of 5S clavams.

5.5 Southern Analysis

The *cvm6*, *cvm6par* and *cvm6/cvm6par* mutants were further characterised by Southern analysis. The results confirmed that in these mutants the chromosomal copies of the relevant genes had been disrupted as expected.

Example 6 – Analysis of *orf3* and *orf3par*

6.1 Construction of *orf3* mutants

Mutants disrupted in *orf3* were made as described in United States Patent No. 6,332,106.

6.2 Construction of *orf3par* mutants

The plasmid p5.7EcoRI ref (pJOE based hyg) was used as the disruption template for *orf3par*. The insert in this plasmid is approximately 5.7kb and includes part of *cvm6par*, all of *orf6par*, *orf4par*, *orf3par* and part of *orf2par* all carried within the plasmid pJOE829 (Kieser, T et al. (2000); Aidoo et al. (1994) Gene. 147:41-6). The disruption vector was constructed by ligation of a thiostrepton resistance cassette (Aidoo et al. *supra*) into *FseI* digested p5.7EcoRI. A unique *FseI* site is located within the insert 507 bp from the start of *orf3par*. The correct construct was obtained and used to sequentially transform *S. lividans* TK24 and then *S. clavuligerus* wild type. Primary transformants were selected on thiostrepton (5 μ g/ml) and hygromycin (25 μ g/ml). The mutants were put through two rounds of sporulation under non-selective conditions as described above and putative double cross-over mutants were isolated.

6.3 Construction of *orf3/orf3par* mutants

The *orf3par* disruption cassette described in section 6.2 was isolated from *S.lividans* TK24 and used to transform *orf3::apra* mutants. Transformants were selected on MYM medium containing thiostrepton (5µg/ml) and hygromycin (25µg/ml). The mutants were put through two rounds of sporulation without selection and double crossover mutants isolated as previously described.

6.4 Fermentation Analysis of *orf3*, *orf3par* and *orf3/orf3par* mutants

To test the effect of disrupting *orf3*, *orf3par* and *orf3/orf3par* on clavulanic acid biosynthesis, spores from each isolate were tested as previously described in section 3.4.

Fermentation Analysis of *orf3* mutants

Mutants Ap3-1, Ap3-2 and Ap3-3 were fermented in Soy medium and compared to wild type *S. clavuligerus* for production of clavulanic acid. After 72hrs growth, accumulations in clavulanic acid were reduced by 31-71%.

From these results it can be concluded that *orf3* is required for efficient production of clavulanic acid as elimination of this gene by disruption causes a reduction in clavulanic acid levels.

Fermentation of *orf3par* mutants

Mutants 3A-1 and 3A-2 were fermented in Soy medium and compared to wild type *S. clavuligerus* for production of clavulanic acid. After 72hrs growth, accumulations in clavulanic acid were reduced by 9%.

From these results it can be concluded that *orf3par* is required for efficient production of clavulanic acid as elimination of this gene by disruption causes a reduction in clavulanic acid levels.

Fermentation of *orf3/orf3par* mutants

Clavulanic acid biosynthesis was completely abolished when mutants 11-1, 11-2, 2-1 and 2-2 defective in both copies of the *orf3* gene were grown in Soy medium and compared to wild type *S. clavuligerus*.

These results demonstrate that under the conditions tested, both genes, *orf3* and *orf3par*, contribute to clavulanic acid biosynthesis as the double disruption results in a mutant unable to make any clavulanic acid.

6.5 Southern Analysis

The *orf3*, *orf3par* and *orf3/orf3par* mutants were further characterised by Southern analysis. The results confirmed that in these mutants the chromosomal copies of the relevant genes had been disrupted as expected.

Example 7 - Analysis of *orf2* and *orf2par*

7.1 Construction of *orf2* mutants

Mutants disrupted in *orf2* were originally made as described in United States Patent No. 6,332,106. These original *orf2* mutants were subjected to a second round of gene replacement to remove the apramycin resistance gene and replace it with a simple frameshift mutation. The plasmid construct used to create the original *orf2* mutant consisted of a 2.1 kb *EcoRI/BglII* fragment of *S. clavuligerus* DNA carried on a pUC119/pIJ486 shuttle vector, with the *orf2* gene disrupted by insertion of an apramycin resistance gene cassette into a centrally located *NotI* site (United States Patent No. 6,332,106). The disruption plasmid construct used in the second round of mutation was derived from the original disruption plasmid by digestion with *NotI* to release the apramycin resistance gene cassette, treatment with the Klenow fragment of DNA polymerase I to fill in the overhanging ends, and then religation to circularize the plasmid. The resulting plasmid construct carries the entire *orf2* gene but with a frameshift introduced at the location of the destroyed *NcoI* site. The construct was used to sequentially transform *S. lividans* TK24 and then the original *S. clavuligerus orf2* mutant. Primary transformants were selected on thiostrepton (5µg/ml) and then subjected to two rounds of sporulation under non-selective conditions. Putative double cross-over mutants were identified based on their loss of apramycin resistance .

7.2 Construction of *orf2par* mutants

orf2par mutants were generated using a PCR-based targeting kit known as REDIRECT (trade Mark of Plant Bioscience Limited, Norwich, U.K.). The plasmids pIJ790 and pIJ773, and the host strain *E. coli* BW25113 were supplied as part of the kit. For this particular application, a pair of oligonucleotide primers, KTA14: 5'-CCATCCCGGCGCCCGTCCGATGCGAAGGAGATCTCCATGATTCCGG-GGATCCGTCGACC-3' and KTA15: 5'-CGGGGCCGGGCATGGTGAACCTCGTCCTCCACGGTGGTTCATGTAGGC-TGGAGCTGCTT-3', designed to disrupt the *orf2par* gene by insertion of an apramycin resistance gene, were synthesized. The *orf2par* disruption cassette was generated by PCR using these two primers with the plasmid pIJ773 as template. PCR conditions used were as described in the user instructions except that no dimethylsulfoxide was used. The *orf2par* disruption cassette was then introduced by electrotransformation into *E. coli* BW25113/pIJ790 which had been previously transformed with the *orf2par* bearing cosmid 14E10 (described hereinabove). Cosmid DNA was isolated from transformants after overnight growth at 37°C to promote loss of the pIJ790 plasmid and analyzed to confirm that the *orf2par* gene had been disrupted. *orf2par* disrupted cosmid DNA was then transferred into

wild type *S. clavuligerus* by conjugation. Conjugation was carried out as described by Kieser, T et al (2000) *supra* except that AS-1 medium (Baltz, R. H. Genetic recombination by protoplast fusion in *Streptomyces*. Dev. Ind. Microbiol 21 (1980) 43-54) supplemented with apramycin at 50 µg/ml was used for recovery of transconjugants. Apramycin resistant *S. clavuligerus* transconjugants were subjected to one round of sporulation under non-selective conditions in order to generate gene replacement mutants as described by Paradkar and Jensen (1995) *supra*.

7.3 Construction of *orf2/orf2par* mutants

The PCR-based targeting procedure used to generate the *orf2par* mutants (section 7.2) was also used to generate *orf2/orf2par* double mutants. In this case the *orf2par* disrupted cosmid DNA was conjugated into the *orf2* mutants described above (section 7.1) rather than into the wild type strain. Apramycin resistant *S. clavuligerus* transconjugants were subjected to one round of sporulation under non-selective conditions in order to obtain unigenomic mutant spores that had undergone gene replacement as previously described.

7.4 Fermentation analysis of *orf2*, *orf2par* and *orf2/orf2par* mutants

To test the effect of disrupting *orf2*, *orf2par* and *orf2/2par* on clavulanic acid biosynthesis, spores from each isolate were tested as previously described in section 3.4.

20 Fermentation Analysis of *orf2* mutants

Mutants defective in the *orf2* gene were fermented in Soy medium and compared to wild type *S. clavuligerus* for production of clavulanic acid. After 72hrs growth, accumulations in clavulanic acid were reduced by 95-98% (Jensen *et al.* (2000) *supra*).

From these results it can be concluded that *orf2* is required for efficient production of clavulanic acid as elimination of this gene by disruption causes a severe reduction in clavulanic acid production.

Fermentation analysis of *orf2par* disruptant

Mutants defective in the *orf2par* gene were fermented in Soy medium and compared to wild type *S. clavuligerus* for production of clavulanic acid. After 72hrs growth, accumulation of clavulanic acid was reduced by 10-30%.

From these results it can be concluded that, like *orf2*, *orf2par* contributes to clavulanic acid biosynthesis as elimination of this gene by disruption causes a reduction in clavulanic acid levels.

Fermentation analysis of *orf2/orf2par* disruptants

Mutants defective in both *orf2* and *orf2par* were fermented in Soy medium and compared to wild type *S. clavuligerus* for production of clavulanic acid. After 72hrs growth, no clavulanic acid production could be detected from the strains contain the *orf2* and *orf2par* mutations.

These results demonstrate that under the conditions tested, both genes, *orf2* and *orf2par*, contribute to clavulanic acid biosynthesis as the double disruption results in a mutant unable to make clavulanic acid.

5 5. Southern Analysis

The *orf2*, *orf2par* and *orf2/2par* mutants were further characterised by Southern analysis. The results confirmed that in these mutants the chromosomal copies of the relevant genes had been disrupted as expected.

10 Example 8 – Analysis of *cvm7* and *cvm7par*

Sequence analysis had identified two additional genes in the paralogue cluster that did not have obvious paralogues in either the clavulanic acid or *cvm* gene clusters. It was of interest to determine if either of these genes was a paralogue to an as yet unidentified *cvm* gene. Therefore the sequence of the *cvm* cluster (WO98/33896) was extended downstream of
15 *cvm3* (*orfup3* in WO98/33896).

8.1 Extension of *cvm* cluster sequence

The cosmid 10D7 (described in WO98/33896) was digested with the restriction endonuclease *SacI*. From this digestion a 6.8 kilobase DNA fragment containing *casI* and
20 *cvm1* was isolated and cloned into a pUC119 based plasmid. The resultant plasmid pCEC019 was used as a template to generate sequence information which allowed completion of the partial *cvm3* gene reported in WO98/33896. In addition, the sequence information showed the presence of another open reading frame, *cvm7*, which was incomplete in this fragment. In order to complete the *cvm7* gene sequence, the next adjacent *SacI* fragment from cosmid
25 10D7, a 1.9 kb fragment, was subcloned. Sequence information was obtained from the end of the clone which contained the remainder of the *cvm7* gene, up to the point where the start codon for the *cvm7* gene could be identified. In total, this resulted in the generation of a further approximately 3.9 kb of new DNA sequence which is described in Sequence ID No.17.

8.2 Sequence analysis

The size of *cvm7* and its orientation relative to the rest of the *cvm* cluster is showed diagrammatically in fig2. Sequence homology searches demonstrated that this gene shares homology with transcriptional regulator genes. In addition *cvm7* also shared 33% identity
35 with one of the two genes identified in the paralogue cluster that did not have any obvious paralogues within the known clavulanic acid or clavam biosynthetic genes. Therefore since *cvm6* and *cvm6par* have been shown to be paralogues, from this sequence data it can be

concluded that *cvm7* and *cvm7para* are paralogues of genes involved in 5S clavam biosynthesis.

Brief description of the figures

- 5 Figure 1. Diagram of the paralogue cluster. The orientation of transcription is shown for each gene (direction of arrow)
- Figure 2. Orientation of *cvm7* in relation to published *cvm* cluster (WO98/33896).
- Figure 3. Annotated sequence of the paralogue cluster

10 Brief description of the sequences

- SEQ ID NO:1 *cvm6para* open reading frame
- SEQ ID NO:2 *cvm7para* open reading fame
- SEQ ID NO:3 *cvm6para* polypeptide
- SEQ ID NO:4 *cvm7para* polypeptide
- 15 SEQ ID NO:5 *cvm6* open reading frame
- SEQ ID NO:6 *cvm7* open reading frame
- SEQ ID NO:7 *cvm1* open reading frame
- SEQ ID NO:8 *cvm2* open reading frame
- SEQ ID NO:9 *cvm3* open reading frame
- 20 SEQ ID NO:10 *cvm4* open reading frame
- SEQ ID NO:11 *cvm5* open reading frame
- SEQ ID NO:12 *orf2para* open reading frame
- SEQ ID NO:13 *orf3para* open reading frame
- SEQ ID NO:14 *orf4para* open reading frame
- 25 SEQ ID NO:15 *orf6para* open reading frame
- SEQ ID NO:16 paralogue cluster
- SEQ ID NO:17 extended *cvm* cluster (underlined sequence denotes new sequence over that disclosed in WO98/33896)
- SEQ ID NO:18 *orf2para* open reading frame (reverse complement)
- 30 SEQ ID NO:19 *orf3para* open reading frame (reverse complement)
- SEQ ID NO:20 *orf4para* open reading frame (reverse complement)
- SEQ ID NO:21 *cvm6* polypeptide
- SEQ ID NO:22 *cvm3* polypeptide
- SEQ ID NO:23 *orf6para* polypeptide
- 35 SEQ ID NO:24 *orf4para* polypeptide
- SEQ ID NO:25 *orf3para* polypeptide
- SEQ ID NO:26 *orf2para* polypeptide

Sequences

SEQ ID NO:1 cvm6para

5 ATGTTCCACCCGGTCTGCCCCGGGGCCGCGAGGACCGCACCGTTCTGGTCTCCGGCCGCGGCTGCACCGTACGGGACAC
 CGAAGGGCGCACCTATCTCGACGCCTCGTCCGTGCTCGGACTGACCCAGATCGGCCATGGACGTGAGGAGATCGCGCAGG
 CCGCCGCGGAGCAGATGCGGACACTCGGTCACTTCCACACTGGGGCACCATCAGCAACGACAAGGCCATCCGACTGGCC
 10 GCGCGCCTCACCGACCTGGCGCCCCAGGGTCTCCAGCGCTCTACTTCACCAGCGGCGGCGGCGAGGGCGTCTGAGATCGC
 CCTGCGCATGGCCGTTACTTCCACCACCGCACCGGCAGCCGGAGCGCACCTGGATCTTGTGCGCGCCGACCGCCTACC
 ACGGCATCGGCTACGGCAGCGGTACGGTGTGCGGGCTCGCCCGCTACCAGGACGGGTTCGGCCCGGTGCTGCCCCATGTG
 CACCACCTACGCGCGCCGACCCGTACCACGCGGAGCTGTACGACGGCGAGGACGTACGGAGTACTGCCTGCGCGAACT
 15 CGCCCGCACCATCGACGAGATCGGCCCGGGCGGATCGCCGCGATGATCGGGGAGCCGGTTCATGGGCGCGGGCGGCGCCG
 TCGTCCCGCGCGGACTACTGGCGCGCGTGC CGCGCTGCTGCGCTCCCACGGCATCCTGCTGATCCTGGACGAGGTC
 GTACCCGCGTTCGGCCGACGGGGACCTGGTTCGCGGCCGAGCACTTCGGGGTGACCCCCGATCTGCTGGTGACCGCGAA
 GGGCATCACTCCGGGTATGTCCCGCACGGGGCGGTGCTCCTGACCGAGGAGGTCGCGGACGCGGTGAACGGGGAGACGG
 20 GGTTCGCGATCGGCTTCACTATACCGGTCACCCCACGGCGTGCGCGTCCGCAATCTCGACATCATCGAACGG
 GAAGGGCTGCTGGAGAACGCGGTGAAGTGGGCGACCACTCGAGCTGGTGTGCGACAAGACGGCCCCGACCCCGTCCCGGGCG
 GGGGACGCTGAGCGGTACGACTTCGCGCTCCAGGAGGCAATCGGCTGGAGCAGCTCCGGTGGGCGCGGTGGAGAC
 25 GCACCTCGGGGTCTGGACGCGCTGCGCGAGGACGCGGGCGTTCATCGTCCGGGCCACGCGCGCTCCCTGGTCTCAAT
 CCGGCGCTCGTGATGGACCGGGCCACGGCGGACGAGGTGGCGGACGGGCTGGACTCGTGCTGCGGCGGCTGGCACCCGA
 CGGGCGGATCGGCGCGGCCCGCGGGGGTGA

SEQ ID NO:2 cvm7para

25 GTGTACGAGTGCAGCGATGAGGTTCTGTCACGACGTCCCGGCCTGCCGGGTCCGTACCGTCCATCACCGTCTCTGGGCTG
 TCTGGGCGTACGCGCCGAGCGGCCGAAACTGGAGCTGGGCCCTCCGCGTCAGCGGGCCGTTTTCGCCCTGCTGCTCATCA
 ACGCGGGCAGTGTGGTGCCGGTCTGACTCGATCGTCTTCCGTATCTGGGGCAACTCACCACCGGGCGCGGTACCGCGACG
 30 CTCCAGTCTATGTGTCCCGGTGCGGAACTCCTGGCCGAGTGTGTGCTCCCGACGGTTCGACACCCGAACTGCTGCA
 CCAGCCGCGGGGTACACCTCGCGCTCGGCACCGAGCACATCGACGCGAACCGTTTGGAGCAGGCATCAGGACAGGGC
 GCCGGCTCTCGCGCAGGAGCAGCACAGGAGGCGCGGCGGTGCTCTGCCAGGCCCTGCTGAGCTGGGGCGGGACACCG
 TACGAGGAGCTGAGCGGTACGACTTCGCGCTCCAGGAGGCAATCGGCTGGAGCAGCTCCGGTGGGCGCGGTGGAGAC
 35 ATGGGCGCACTGCTGTCTGCGGCTGGGGCGGACGAGGAGGTGATGGACCAGTCAAGCCGAGGTCAGCGCAATCCGC
 TGCGGGAGCGGCTGATCGGGCAGCTCATGACGGCGAGTACCGGCTGGGGTGCCAGGCGGACGCGCTCAGGACGTACGAG
 GCGACGCGGCGGGCCCTGGCCGAGGAGCTGGGGACCGATCCGGGCAAGGAGCTGGCGGCGCTGCACGCGGCGATCCTGCG
 TCAGGACAACGGTCTGGACCGCGTCTGCCCGCGTCCGCGCGCCGTCCGCGGGGTCGGGCGGGGGCGGCTGACGGTGT
 CGGTCCCGGCACAGCGGTGAGGCCGTTGACGCGCGCGGTGGCGGGGCGGGCGGGTCCCGGGGGCGATGACGGTGGCG
 40 GCGGGCGCGGGGGCGGCCCGCGTCCGCTCCGGCTCCGTTTCCGCGTCCGTTTCCGCTCCGGCTCCGGCTCCGGCTC
 CGTCTCTGCGTCCGTTCCACCTTCTTTCCCGGTCCGTTTCTGGTCCGCGTCCGTTGCCGCTCCGTTAGCCGCGCCCCG
 TTTCCGGCCATGTCTCCGGGCCCGGGTCCGTTTCCGGTCCGTTGGCGCTCCACCGGCCGAGACCTCCGGGGCGAGCCG
 45 GTCCACGGGGCGCGCAGGGGATGCGCACCGGCGAGTGTTCCACAGCTGCCGCGGTTCCGCGCGCGGCGAGGACT
 GCGCGGTCTGCTGAGTCCGCGACGTCCGCGTCCACACTCGGGCGGGTGGCGTTCTGCTCGTCCGCGAGGCGGGCAGCG
 GCAAGACCCGCTCTCTCCGAGTTGGAGCGCTCGGTTCCGGACAGTGTGCGCACCGTCTGGCGCTCCTGTTCCGAGAGT
 50 GAGGACCGGCCGACTACTGGCCGTGGACGACCGTGTGCGGCATCTGTACGCGATGTGGCCGGAACGTATGCACGGATT
 CCCCAGTTGGTTCGCGCGCGCACTCGCGGAACTGCTTCCGAGGTGGGCCCCGAGCCACAGGGGCCGCACTCCCCGAGC
 GGGGCGAGGAGAACAGCGGCAACGGGGACGGTTCGGGCGACGGGACAGCACCCCGCGCACACCTTCACGCTCGCGCCC
 GCTCTCGCGCCCGCGCTCCAGAGAGGCTCGTTTACCCTGCACGACGCGGTGTGCCAGGCGCTTCTGCGCACGGTCCG
 CGAACCCGTGGTATCATGCTGGAGGACATGGAGCGGGCGACGCCCCCTCGCTCGCCCTGCTGCGCTCCTGGTGGAGC
 55 AACTGCGCACCGTCCCCCTGCTGCTCGTGGTACCACGCGCACCTTCCGGTCTCGCGCACGACGCCGAGCTGCGACGGGCC
 GCCGCGGTGATCCTCAGTTCAGCCGCGCGCGCGGGTCTGCTGAACGCCCTGGACGACGGGCCACCGGGAACTCCG
 CGAGGGATGCTGGGCAAGGCCCGGACACCTCTCGTACGGGCCCTGCACGAGCGCTCCGCGGGAACTCCGTA
 TCGTCCAGCTCTCCGCTCGCTCCGCGAGGGCTCGCGCCGGTGTGCTGACATCTGCGCGGTCTGAGAGCGAGTTGCGA
 60 GTGCTGCAACGGCTGTGAGCGTGC CGCCCGCGTGC CGCGGGTGTGCTGACATCTGCGCGGTCTGAGAGCGAGTTGCGA
 ACGGCGTGTGATCGAGACCGTGTGCGCCATGAGGGAATCCCGTGGAGAACGTCGTAACGGCGGTCCGCGCGGTCTGC
 TGGAGGAAGACCCCGACGACCCCGGGCGGCTGAGGTTCTGTGATCCGCTGGTCCGGGAGGCCGTCTGGACGACCTGGAG
 AACACCGTCCGCGCGTSTCVMARGTCCGTTCTCCGCGTCCGGGCGCTGGCCACGGTCTGA

SEQ ID NO:3 cvm6para polypeptide

55 MFHPVLPRGREDRTVLVSGRGCTVRDTEGRTYLDASSVLGLTQIGHGREEIAQAAAEQMRTLGHFHTWGTISNDKAIRLA
 ARLTDLAPQGLQRVYFTSGGGEGVEIALRMARYFHRTGSPERTWILSRRTAYHIGYGSSTVSGSPAYQDGFPGVLPVPHV
 HHLTPDPYHAELYDGEDVTEYCLRELARTIDEIGPRIAAMIGPEPVMGAGGAVVPPDYWPVRAALLRSHGILLIIDEV
 VTAFGRGTWFAAEHFGVTPDLLVTAKGITSYVPHGAVLLTEEVADAVNGETGFPIGFTYTGHPACAVALANLDIER
 60 EGLLENNAVKGVDHLAAGRLAALRGLPAVGDVRLGMLLAVELVSDKTARTPLPGGTLGVVDALREDAGVIVRATPRSLVLN
 PALVMDRATADEVADGLDSVLRRLAPDGRIGAAPRRG

SEQ ID NO:4 cvm7para polypeptide

65 VYECSDVEVRHDPGLPGSPSITVLGCLGVRADGRKLELGPFRQRAVFALLLINAGSVVPVDSIVFRIWGNPSPGAVTAT
 LQSYVSRLRLKLLAEVLPDGTPELHQPPTYTLALGTEHIDANRFEQAI RTGRRLSREEQHQEAREAVLCQALLSWGGTP
 YEELSAYDFAVQEANRLEQLRLGAVETWAHCCRLRLGRDEEVMQKPEVQRNPLRERLIGQLMQAQLRGCQADALRTYE
 ATRRALAEELGTDPGKELAAALHAILRQDNLDRVVPASAPPSAGVGRGAVTVSVPAQRSRPLTRPVAGRARVPGAMTVA

AGAGAAPASASGSVSASVSGSGSGSAPASVPTFFPGSVSGSASVAASVAAPVSGHVSGPSAFGSVALHRPQTLRGEP
 VHGAQGMRTGQVFTLPFPVGRGDELRLLESATSAFHTSGRVAFFVVEAGSGKTRLLSELERSVPDSVRTVWASCSES
 EDPRDYWPWTTVLRLHYAMWPERMHGFPGLRRLAELLPEVGPEPQGHSPDGGEENSGNDGAGDGDSTPAHTLTLP
 ALAPPRSREARFTLHDAVCQALLRTVREPVMIMLEDMERADAPSLALLRLIVEQLRTVPLLLVVTTRTFRLAHDAELRRA
 AAVILQSTGARRVLLNALDARATGELAGGMLGKAPDPTLLVRLHERSAGNPYFLVQLRLSLRQGLAAWETEI PDELAGV
 VLQRLSSVPPAVRRVLDICAVVERSCERRVIETVLRHEGIPLNVRTAVRGGLLEEDPDDPGRLLRFVHPLVREAVWDDLE
 NTRRPVSRSSALGALATV

SEQ ID NO:5 cvm6

GTGCCCGGCTCCGACTCGAAGCACTGGACCGTGCCACCCTCATCCACCCACCCTCTCCGGAACACCGCGGAACGGAT
 CGTGCTGACCTCGGGTCCGGCAGCCGGTCCGCGACACCGACGGCCGGAGTACCTGGACCGCAGCGCGTCTCGGGG
 TGACCCAGGTGGCCACGGCCGGGCGGAGCTGGCCCGGTCGCGGCGAGCAGATGGCCCGGCTGGAGTACTTCCACACC
 TGGGGGACGATCAGCAACGACCGGGCGGTGGAGCTGGCGGCACGGCTGGTGGGGCTGAGCCGGAGCCGCTGACCCGGT
 CTACTTCACAGCGGCGGGGCCGAGGGCAACGAGATCGCCCTGCGGATGGCCCGGCTCTACCAACACCGGCGCGGGAGT
 CCGCCCGTACTTGGATACTCTCCCGCGGTTCGGCTTACCACGGCGTCGGATACGGCAGCGGCGGCGTCAACGGCTTCCCC
 GCCTACCACAGGGCTTCGGCCCCCTCCCTCCCGGACGTGCACTTCTGACCCCGCGCAGCCCTACCGCCGGAGCTGTT
 CGCCGCTTCGAGCTCACCGACTTCTGCTCGCCGAACCTGCGAGACCATCGACCGGATCGGCCCGGAGCGGATCGCGG
 CGATGATCGCGCAGCCGATCATGGGCGCGGTTCGGCGCCGCGGCCCGCCCGGACTACTGGCCCCGGGTTCGCGAGCTG
 CTGCACTCCTACGGCATCTGCTGATCTCGACGAGGTGATCAGGGGTACGGGCGCACCGGGCACTGGTTTCGCCGCCGA
 CCACTTCGGCGTGGTCCCGGACATCATGGTCACCGCCAGGGCATTACCTCGGGGTATGTGCCGCACGGCGCCGTCCTG
 ACCACCGAGGCGGTTCGGCGACGAGGTCTGCGCGACGAGGCTTCCCGCGGGCTTCACTACAGCGGCGCATGCCACGGC
 CTGCGCGGTGGCTGGCCCAACCTGGACATCATCGAGCGCAGAGAATCTGCTCGACAACGCCAGCACCGTTCGGCGCTTACC
 TGGGCAACCGCTGGCGGAGCTGAGCGATCTGCGGATCTGCGGGACGTCCGGCAGACCGGTCTGATGCTCGGTGTGAA
 CTGGTTCGCCGACCGCGGAACCCGGGAGCGCTGCGGGCGCGCCGTCGCCGAGGCCCTGCGCGAGCGGGCGGGCATCT
 GCTGCGCGCAACCGCAACGCCCTCATCGTCAACCCCCGCTGATCTTACCCAGGAAGACGCCGACGAACTCGTGGCGG
 GCCTGCGCTCCGTACTCGCCCGCACCGGCGGACGGCCGGGTGCTCTGA

SEQ ID NO:6 cvm7

ATGAAGTACGACATAACCCACCATCCGGCCTTCGGTTCGACCTCTCGGCCCGTTGACCGTGACCGCGGCGAGCAACC
 CGTGGACCTGGGCGCGCCACGGCAGCGCGCCCTGCTCGCCCTGCTGCTCATCGATGTGGCAACGTGGTCCCGCTGCCGG
 TCATGACCGCGTCGATCTGGGGGGCCGACCCACCGTCCCGGGTCCGGGGGACGCTCCAGGCTTATGTGTCCCGATGCGG
 AAACCTCTGCACCGCCATGACCGTTCCTTTCGCTTGTCCACAGCTCCAGGGGTATCTCTCGAAGTGGATTTCGGCGAA
 GGTGGACGCGGTGGTTTTTCGAGACAGTGTTCAGGCGTCCGGGAATTGAGCAGGGCCCGGAACCCGAGGCCACCCGGG
 CCGTGGCTTGGCTCGCCCTGGAGATGTGGAAGGCGACACCCATGGGCGAGTGCATGATTATGAATTTGTGGCGCGGAG
 GCCGACCGGCTGGAAGGAATCCGGTTACGCGCGTGGAGACCTGGTCCAGGCGTGTCTCGATCTCCAGCACTATGAAGA
 GGTGCTATTCAGCTCGGCGAGGAGATCCACCGCAATCCGGAACCTGGAACGGTGGGCGGTCTCTTCATGCGGGCCAGT
 ATCATTCCGGACGGTTCGGCGGAAGCCCTGTTGACGTATGAACGTATGCGTACCGCGGTGGCGGAGAATCTGGGGCCGAT
 ATCAGTCCGGAGCTCCAGGAACCTCATGGAAGATTCTGCGCAGGAACCTCACGGAGACACCCGCGCGCATCGACGGC
 CTCCCTCACACGGGCGGCGGGCCCGCACGGGCCCCCGCCCTGGCGGAACCCGCGACCCCGCGGACCCGCGGACATGG
 CCGAAACACGGTGGCGGAGGAAGCGCGCGCCCCCGCCCCGGCGCGCCCGGACCCCGCCCCCATGCCGTCCCCC
 GTACCGCTCCCCATCCGTACGGGGCGCTCCCGCGGTCAACCCGCTGCTTCCCCCGGTCCCCCGCTCGGCCCTCCGTT
 AGCGGCACCCCGCGAGACCGAGGACCCGGAACCGGCGCGCCCCCTCCCCCTCCGCGGGCGGCGGACTCATCGGCCG
 GCGCGAAGTGCAGGCTGCGGCTGCTGCTGACGAGACCCGCGCGGGCCACGGCCATGTCCTGCTGGTCTGCGCGGA
 CAGGCGATCGGGAAGACCCGGCTCCTGGAGCACACCGAGCACACCTGGCCGCGGGCGCGTTCGGGTGGTTCGTTTCGCA
 CTGCGTTCGCCACCTCCCGGCACCGGGCTACTGGCCCTGGGAGCACCTCGTACGCCAGCTCGACCCGACAGCGGCCTCG
 GTGACGACGGCGACCGCGACCCCGTCCGCCAGGCGAGTGGCTGCGCGGAACACCACTACCCACAGATGCGGATCTGC
 CGGACGGTGTCTCGCGCGGCGCGGCGGACCCCGTCTGTTGATCCTGGAGGATCTGCACCTCGCCACCGCGCGGTCT
 GGATGTGCTCCAGCTCCTGGTCAAACAGATCGGCGAGGCCCGCTCATGGTCTGCGCCACCTGCGCGAGCAGATCTCG
 CCGGGACCCCGCGTCCGCGGGCGCTGGGCGCATCTCCAGGCGGGCAACACCGGCACCCCTCCGCTGGACGGGCTC
 ACCGAGGAGCAGAGCCGGGAGCTGATCGTCTCGGTTCGCGGGGGCCCCGTTCCGCGCCCATGACGCCAACCGCTCCAGCG
 CGCCTCGGGCGGAACCCGTTTCTGCTGCTCAGCATGGTACAGGGGAGGACGGCACCCAGGAGTGGGACCGGCGGTGCG
 TCCCGTTTCGAGGTGCGCGAGGTGCTGCACGAGCGGCTGAGCGAATGCTCCCCGTCCACCCAGGACGTGCTCAGCTCTGC
 GCCGTGCTCGGCATGAGCGTGCGCCGACCGTGTCTACCGACATCATGTCCACGCTCGACATCCCGCACACCGCGCTCGA
 CGACGCGCTCGGCACGGGGTGTGTCGCCACGACCGGAACACCGACGGAATGGTCCACTTCGCCCATGGGCTGACCCGGG
 ACTTCTGCTCGACGACACCCCGCGGTACCCCGCGCCGCTGGCACACCGGTTCGCCGCCACCCCTCGCCCTGCGCTTC
 CAGCAGGGCGACGACACCGCGAGATCCGCCGCACTGTCTGGCCGCGGCCCGTCTGCTCGGCGCCCGCGCGGGGTGCG
 CCCCCTGCTGGCGCTGGCCGACCGGGAGCAGTCCGCTTCTCCACGCGGAGGCGTTCGCTGGCTGGAGAGCGCGGTTCG
 CGGTGCTCGCGGCGTGCCTCCGGGACCGCGGTGTCCGCGTCAACTCCAGTTGCGCAACCGATGATGGCGCTGCAC
 GCGCTGATGACCGCTATGGATCGGCCCGCGTCGAGACGTTCTCTCCAGGTACCCAGTGGGAACAGCTCTTCGACAA
 CACCCAGCCACCGGGCTGCTGCACGTCCAGGCGCTGAGCGCGTCAACCGGGCGCCATGAGCAGGCGGCGGAGCTGG
 CCGGGCTGCTGCACGAGCTGGCCGACACCGGCGGCGACCGGAGGCCCGTTCGCGCGCTGCTATGTGACGCGGTCAAC
 CTGTATGTGGGCGGACGGTTCGACGAAGCCCTCGCCGCGTTCGCCAGGGCACCGAGATCACGGACGCCCTCTGGCCGG
 ACACCGCAGGACCGCCGCCCGCACCGGCGGCGGCGACCTCCAGGACCGGCTATCGACTTCCGCGCTATCTGGCGCTCG
 GCCACTGCTCTCAGCGCGACCGGATTACAGCCAGCGCTACCGGAGGAACTCTCCACCTACCCAGTCGGAACGGTAC
 GACCGCGCTCGGACCGGGCTTCGCCCGCTATGTGGACGCGTCTATCGCGTACGGAGTGCATGTCCAGGGGGTGTG
 GCTGGCGCGCGGGCGGGGCTCGACCTCGCCCGCGTGCAGCTCCCGTCTGGCAGCGGATGCTCGCCGTCCCCCTCG
 GCTGGGCGGAGTCCACAGGGGGCGCACGACAAGGGGTGGCCCGGATGCGGGAGGCGTGCACGAGGCGGCGCGGCAC

CGGACCCTGCTGCGCCGTACGCTCCACCTCGGCCTGCTCGCCGACGCCCTCCAGTACACGGGCGCCCGGGAACAGGCCCG
GCGCACGATGTCTCCGCCGTACGGGAGATCGAGCGCCGCGGCGAGTACTTCTGTCTCCGGCCGAGTGGCCCTGGGCCC
GGCTCTCCACAGCCACGGCACCTCCGCCGCGGCGGAGCACGGGTCTGTCCACGGCAGGCACTGA

5 SEQ ID NO:7 cvm1

ATGTCCCGCTCTCCGCCGAGTCCCCGGCCGGTTCCGTGTCCGCCGCGGTTCCGCGTCCGCCGGTCCGCGCCCTCGGGGA
CCTTCCGGTCACTGCCAGGGGCTCGGCTGCTGCCGACCACCGACTTCTACGGACGCCCGGACCGCGCCCGGGCGACGG
CCACCATCCGCGCCGCCGTGACGCGCGGGGTACCCCTGCTGGACACCGCCGACGTCCAGGGGCTCGGCGCCGGTGAGGAG
CTGCTCGGACGGGCGGTGCGCGGGCCGCCGGGACGAGGTGCTGATCGCCACCAAGTTCGGCATGGTGCCTCGTCCGACGG
10 CGCCTCCCAGGGCTTGTGCGGCGAGCCGTCTACGTCCGCGCGGCTGCGAACGGTCCCTGCGTCTGCGCACCGACC
GCATCGACCTGTACTACCAGCACTGGACGACCCGGCGGTCCGATCGAGGAGACCGTGGGTGCGGTGGCCGAGCTGGTG
CGCGAGGGCAAGGTCCGACGGCTCTCTCCGAGCCCTCCGCGGCCACGCTGCGCCGGGCGGACCGGGTGACCCGGT
GACGGCGGTGACAGCGAGTGGAGCCTGTGGTTCGCGCGGGATCGAGGACGAGGTGGTCCCGTCTGCGGGGAGCTGGGGA
TCGGGATCGTCTTACGCCCTCTGGGACGGGTCTTCTACCGGCACCATCCGACACCGACGATCTGGGGGACGAG
15 GACTTCCGCCGGGGCCAGCCCCGGTTACGCGTCCGGCCCTCGCGCGCAACCGTCTGTGCTGACCCGGTTCGCGCCGGT
CGCGGACGGTCTGGGGTACCCCTGGCACAGCTCGCGTCTGCTGGTGCACACCGGGGCGAGGACGTCTGCCGATCC
CGGCGACCGCAACCCGGCCCATCTCGCGGACAATCTCGCCGCGCCTCGATCCGGCTGGACGACCGGTCCCTCGCGGAG
GTGACGGCCGCGATCTCCACCCGGTGTCCGGGGAGCGGTACACCCCGCATTTGCTCGCCATGATCGGCAACTGA

20 SEQ ID NO:8 cvm2

ATGTCCGTGGCATCGGCCGGTATGACGGACGAGCAGCGCAAGGCGGTCTACCCGCTACTTCAAGGCGTTTCGACAACGG
CGGCGTCCGACGACGACGGCACCCCGCGATCGACTACTTTCGCCGAGGACCGGCTCTTCTTCTTCCCAAGTGGGGTCTGG
CCCGGGCAAGTCCGAGATCGCCCGGCTCTTCGACGACCTCGGGGGACCATCCGCTCGATCACCCACCATCTGTGGTCC
GTCAACTGGATTCTGACCGGGACCGAACTCTCGCCGCGGAGGGACACCCACGGTGAGACCGGGACGGGCGTGGCG
25 GCGGGTGACCCGAGTGGGCGCCCGGGCGCTGGTGCACGGTCTACGAGGTGCGGGACTTCTCTGCTCCACCGGGCCTTCG
TCTATCTGGACCCCGATTACGCGGGCAAGGACACCGCGCTTACCCGTGGCTGTGA

SEQ ID NO:9 cvm3

GTGACCCGGCCTCCGGGCTTTCCGCGCACACCCACGGGTCCGTGTCCGGGAGTCTGCTGCGCCGGGTGGCGGGCCACTA
30 TCCACCCGGGTGGTCTGTGTCACCGGTCCGGCCGAGGCTCCGGGGCAGCCGCGCCCGCCATGGTGGTGGGGACGTTCA
CCTCGGTGTGCTCGATCCGGTGTGTTGGTGGGTTTCTCCCGGCCAGGTCTGTCGACGACCTGGCCGCGGCTCCGGGCGGCC
GGGCGTTTCTGCGTCAATGTGCTCGGCGCGGATCAGGGCCCGGTCTGCCGAGTTTCGCCGGGGGCGATCCGGGGCGCTG
GGAGGTGCCGTACCGGACGACGGCCACCGGTCCCCCGTCTGCTCGACGCGCTCGCGTGGTTCGACTGCGAGGTGGCGG
GGGAGACCGGAGGCGGGCGACCACTGGTTCTGTCACCGGGCGGTGCGCGACCTCGGGGTGATCCGCGAGGTTTCGCCCTG
35 GTCTTCTGCGGGCGACTACGGGCACTGGGCCGGGGCGGCGGCTCGGGCCGGGCGGGGCGGCGGTCCGCCGTCTGCC
GGTCTGA

SEQ ID NO:10 cvm4

GTGGAATGCCGCATATTTCGAGATCGACGAACTGCCGTTGCTGGACGGGGAGGTCTGCGGGACGCCCGGATCGGTTACGC
40 CATGTACGGCACGCCGAACGCCGACGGGACGAACTGGTGTCTGTCCGTCTTCTCGGCCGGGACCAACCGGGTACG
ACTGGCTGATCGGTGCGGGGCTGCCGTGGACACCCGGCGGTACTGCGTCTGTCACCGCCGGAATCTTCGGCAACGGGGT
TCCAGCTCGCCCGCAACCAACCGTCCGGGTCCCGCTTTCGCTGATCACTCCGACGACAATGTGCGGGCGCAGCACCG
GCTGCTACCGAGGAGCTGGGGGTACGGAACTGGCCCTGGTACGGGCTGGTTCGATGGGCGCGGGCCACGCTACCACT
GGGCCGTGTGCGATCCGGGGATGGTGCGCCGGATCGCCCCGATCTGCGGGGCGCCGGTGAGCAGCCCGCACAGCTGGT
45 CTGCTGTCCGGTCTGGCCGCGGCGCTCAGCGCCGACGCGGGGAGCGGGGCGGAAGGCGGCGGGCGGGGTGTTTCGCCG
GTGGGGACCTCGCGTTCTTCTGGGCCCGCGTGCACACGGGAGTGGGTTTCGCCACCCGCGACATCATGGGGCAGG
GCTTCTGGGAGCAGGTCTTCTCTCGGGCCCGCGCGCGCGGATCTGCTACCATGGTGCACCTGGGAGAACACGGAT
GTGGGGGCGACACCCGGGGCGGGGGGAGCGTCAAGGCGGCGCTGGCTCCGTACGGCGCGGGCCGTGGTGTGCGCGG
CGCCCTGGACGTGTGTTTCGCCGTGAGGACGAGAAGCGGTGGCCGATCTGCTGCCGTATGCTCTGCTGGAGGTGATCC
50 CGGGAGTGTGGGGGATCTCGCGGGGTCCGGGGGTGCGGCCCGACCGGGAGTTTCGCGGGGCGCGCTGCGGCGGCTG
CTGGACAGCCCGGTGGACGGGGGCTGA

SEQ ID NO:11 cvm5

GTGAAGTCCATTCTTCTATCTGCCAACGGTCCGGCAGTCAAGCGAGGTCCAGCGGGGTATGGCGGGGTCAATCCGCA
55 GAACTACCAGAACATGCTCCGGCAGCTACCCGGCAGGCGCAGGCGGCCGACGAACTCGGCTACTGGGGACTGTCTTCA
CCGAGCACTTCCACACCGAGGTTTCGAGGTCTCAACAACCCGATCATGCTGGGGCTCTACCTCGGCATGCAGACC
CGGCACATCCGGGTCCGCCAGATGGCCAACGTCTGCCGTGCACAATCCGCTGCGGCTGGCCGAGGATCTGGCGATGCT
CGACCACATGACCCGGGGCGCGCTTCTGTCGGGATCGCGCGGGTTCGAGAAGCGTGGGCGGACATCATGGGGCAGG
TGTACGGGTTCGGCGGCACCTGTCCGACCGGGGAGCGGGACCGGCGCAATCGTGCCCTCTTCGAGGAGCACTGGGAG
60 ATCATCAAGAAGCGTGGACGACCGAGACGTTACCCACTCCGGGGAGCAGTGGACGATCCCGGTGCCGACCTGGAGTT
CCCCTACGAGGCGGTGCGCGCTACGCGGGGGCCTCGACGAGAACCGCGTCAACCGGAGGTGGGCATCGCGCCCAAGC
CCTACCAGCGCCCCACCCGCCGTCTTCCAGCCGTTCACTTCACTGAGGACAGTTCGGTTCTGTGCCCGGAGGGC
GTGGTGGCGATCTGTGAACACCGACGACCATGCTCGCCCGGCTGATGGACATCTACCGGGAGGAGGCGGCGGCG
GGGCCACCGTCTGCGCGGGGCGAGCGGGTCCGGGTGATGAAGGACGTTCTGGTCTCCCGGAGCTCCGGCGAGGCGG
65 ACCACTGGGCGTCCCGCGGCGGCGGCTTCATCTTCGAGAAGTGGTTCGGCCCATGGGCTTACCGAGGCGCTGCGCGG
ACCGGCGAGACGGTCCGATCGGCTCGGACTACAAGACCCTGGTGCACCGGGGGTGGAGTGGGTGCGCACCCCGGACGA

CATCAACCGCATGATCGAGAAGCTGGTGGAGCGGCACGATCCGGAGTATCTGCTCCAGTGCCAGTACTCCGGGCTGATCC
CGCACGATGTCCAGCTGCGCAGCCTGGAGCTGTGGGCCACCGAGATCGCCCCAACTGGCTCTGA

SEQ ID NO:12 orf2para

5 TCAGATGGCCAGGGCGGCGAAACCGCCGGAAGTCTAGGCCACCGGTACCTCGATCAGGAACGGGCGGCCGAGTC
CGGCGCCCTTGGTGAGGGCGGCGAGCAGCGAGGTGCGGTGCGGTGGCGCGGACGGCCTCGCAGCCGTTGGCCTCGGCGAGC
TGGACGAAGTCGACGCTTCCGAAGCCGACGGCGGGGGCGTGGGAGCGTGGTGTCCGAGGTTCTGGTACAGCTCGATCAG
GCCGTTGCGGTGCTTGTGACGACGACCATGACGATCGGCAGGCCAGGCGCACGGCCGTCTCGATGTGCGCGCTGTGG
10 AGTGAAGCCGCCGTGCGCCGCGATGAGGAAGACGGGCTCGCCGGGCCGGGCGATCTGGGCGGCCATGGCGGCGGGCAGT
CCGTAGCCGAAGCTGGAGCAGCCCGCGGAGGTGAGGAATCCGTACGGCTGGTGGACTTGGCGAAGAGCACGCCGTAGTG
GCGGAAGAAGCCGATGTGCTGACGAAGGTGCCGTTTTCGAGGACGGAGTTCATGCACTGATCACCTGGTGGACCCGCA
TGCCGTCCTCGTACTCGGTGGGGTGGCGAGGAATTCGGCGACGCGGGCGCGCAGGGCGCTGAGGTGTCGCCGGGTCTTG
GGGGCGAGGCCGAGGTGCGCTGTCGAGCGCGGTGACGAATTTCGGCGACGTTGGTGACGATGTGATGTGCGCGCGGAA
15 CAGCTCCGGGATCGGGTTGACCTCGGGGGCGACCCGACCGTGGTCTTGCGCCCGGCCCGCGTCCACATGGAGGGGCGCA
GGTCTCGGCTAGTCTGACCGATCGCCAGGAGGAGGTGCGCGGGGCCGAAGATCTCGTCGAGGGCCGGGTGGCCGAGA
ATGCCGTCCATGTAGCCGTGATGCGCGCGTGTGAGCGGGTGGTCTGTCGCGCAGGACGCCCTTGCCGCTGATGGTGGT
GACGACGGGATGTTTACGCCGTGCGCGAGGGCGCGCAGGGCGTCGACGGCCCGCGCGGATGACGGCGCTACCGACGA
CGAGGAGGGGGTCTCGGCCTCGCGCACCAGCTCAGCGCCCTCGTCGAGGCGGGCGCGCCAGTCGGCGTCCAGGGCGTGG
20 GTGGCGGTGGCCCGACAGGGGGCGTGGTGGGGTGCCGTTTACGCTCGCGCGCGAGGAGGTGACCGGCGAGGCTGAT
GAAGCTGGGACCCACGGGCTCGATCCGGCTGTTGAGGACGGCGCTGTCGACGAGGTTGACGATGTCTCGCCGCGTTTCA
GCTGGACGCTGAACCTTGGTCAGCGGGCCCATCAGCGCGGTGCTGTCCAGGCACTGGTGGGTGACGTTGGGGTAGCAGTCG
TACGACTCGGACTGCGCGGCCAGCGCATGACCGAGCTGCGGTCCAGGGCGAGGTGGCGACGCCGGTGGCCAGGTTGGT
CATGCCGGGGCCAGGGTGCAGAACGCGCTGGGGGCGGTTGGTGATCCGGGCGAGGACGTCCGCCATCACCCCGGCGG
25 TGAACCTCGTCCGGGTGAGGACGAAGTCGAGTCTTCGACCTCGTCGAAGAGAATGGCGGACGCCCTCCCGGCGGACGACG
CCGAATACATGGTCGACACCGTACTGGTGAAGACGTTCCAGCATGGCTTTCGCGGTGCTGGTGGCCAT

SEQ ID NO:13 orf3para

TCATACGACCACCCGGCCCTGGAGCCTGAGCCTGCGCACCGCGTCGACGGAGCGCCGACCGTCTCGCCGAAGTCCACGT
CCTCCGGCGGCACCGTGTGATGACCACCGCGTGTACAGGCGCCGTGCCATGGCGCCCTTGACGGCCGTACCTCGTGC
30 CGCCGGATCCCTTCGCGGAGGAGCAGTCCGGTCCACGCGCTGGTGGTGCCGAGCCCTCGTGGATGCCAGCTTGGGGCG
GGCCACGGTCTCGGCGGGCAGCAGGCCGAGAGGGCTGCCGAACACCCACTTGTCGGTGCCCCCGCGCGGCTTTGAGCC
CGGGTTTCGAGGAGACGCGCGTCCAGGACCGCGCGGTGCTGTCCAGGCACTGGTGGGTGACGTTCCCGGCGATGCCCGCG
AGGACGGGGGACATCTCGTTGAGGCCGTGAAGCCCGCATGTGCGCCGCGATCTCGTCGTCAGGGACAGAGCGAGGC
35 CGTGCGCCGGTGATACCGCCGAGCGGGATGTGCGCGCCGTACCCGGTGAGGATGCGGAGCGGCCCGGTGTGAGCGGCC
GGTAGAGGGCGACGAGCGGCGAGCAGGTACTCCAGGACCGTGGGGTGGGTGATCTCGCGCGCGGACCGCCAGGGCAGT
TCCCTGACGAGTTCGGCCGAGTGGAGCCGATCTCGTGTGCGCGGTGCCAGGTGGACGGCGACCGAGCGGGCCGCGTC
GAACTCGTCGGACACCTCGGTGCCATCGACACGACCGGTGTCGGGGTGCCAGGGCCCGGTGTGGGCGGCGACTCCCC
CGGATCGATGCCCGGACAGGACGAGCGTGGGGGGCGCCCTCCCGCGCGCAGCCGGGTGCGGACCGCCGTGGCGAGG
40 CGTTCGCGGACAGGTCCACCGCCTCCCGTTGCGCGGGCAGCGCCGGGAGAGCGGGGGTGTCCAGGTGCGGACCGCCCT
GGCGGTGATGTGCGAGCGCGGACTCCGTGACGAGGAGGGCGGTCCCGGCGGGGACCCGGCAGACGCCCGCCCGCCCG
GCGCGGTGTGGGTGCGGACAGGCCAGCGCCGGCCCGGCTCGTGCGCCAGGGTCTTCGCTCGGTGGCGGCGCTCAGC
CCGTCACGTCCGGCGGACGACCGGTACCGAACCAGCGGTGGTGGTGGCCGCGACGCTCGCGCGGTGGAGGCGTTC
45 GGTGAGCAGTGGCGGAACCGTCCGTTTACGAGCGGAGGCGCCCGGGGCCAGCGCCGCGGACGCGGCGGAGGAGC
CGGCGTTCGCGGAGGGCGGAGAGGAGCCGCGGAGCGCTCCGGTCAGCTCGGCGCGGTTGTACAGCTCGCCCGCAGGAGC
AGCCGACCTGGCCGTGCGCGACAGGACGGGCGGACGCGCCAGGGTCACGGCCGTTCCGCTCCAGAGCGGGTACGCGGT
GCCGTGTCGACGGGACATGGGTCCCGCGGACGGCGAAGCGGGTGCGCTGCCGGGTTTCGGAGTGACCGCGGGGCGCG
CGCCGGGGCGGCCCTCGGTGCCGATGCGCACCCGGAATCCGTACACGAGGTGCGGGCCGGCAT

SEQ ID NO:14 orf4para

CTACCCCCACCGTGCCCGGCGAAGTCCACGGCGCTCTCGGCGTCCACCGGTCCACCGGTTCTCGGCGTTCTCGGCGT
CGTCCGCGCGCCCGCCCGGTGGCAGGGGAGAGTCCACCGGTGCGGACGCGGGCGACGTGGTGGCGCGGGCGTACTGGTAG
AGCAGTTCGCGCCCGATCTCCGCGCCAGCAGGGAGGTGATCCCGACGGGTGTCACGCGGGGACACCTCGACCACGTC
GAAGCCGACGGGCGTGGCTGCCGACACGTGAGCAGGGTCAGCACCTCGCGCGAGGACAGCCCGCGGGGGCCGGTG
50 TGCCGTTGCCCGGGCGTACGCCGGGTGACGACGTCGATGTGCGACGGAGACGTACAGCGGACGGCCCGCGACGGTGGCG
CGGATCTGCTCGGCGATGCCGCGCGGTGAGCGCCGGGTGAAGTCGGCGCGGTCAGATGCTGACGCGGTGCCCGCGCGC
GTAGTCCAGGAGTGGGGCCGCGATTGTGGCCGCGGATGCCGACCTGGACAGGCGCTCCGGGTCCACAGGCCCTCTT
CGATGGCCACGCGGAAGGGGGTCCCGTGGTGGTAGGTGCCCGGTGACGCGGTGGGTGGTGTGCTGTGCGGTCCAGG
TGCAGGACGGCGACCCGGCGGTGGCGGGCGTGCACGGCGCGCAGGGCGGCCAGGGAGAGCGAGTGGTCCCGCCGACGAT
55 CAGGAACGCGTGGTTCAGGAGCCGGTCCAGGAGCCGGTACGGCGACCGTCGCGGTGTCATCGCCAGGCGCGGAGGCG
TGAGGTGATGTGCGCCCGTTCGACCAGTTCGATCCGGTGAAGACCCCTGGGCCCCGGTTCGATGCCGACGCGGTGGATC
AGGCTGGACTCGTGGCGATGGCGCGGGCGGACCGCGCGCGGGCCGGTAGCTGGTGCCTCCGTGTCAGGGGCGCC
GACGACACCACTCATGGCCGATCGGGTGGGGCGGTGGCGCAGCCGATGAAGGTGCGCGGTTGGGCGTAGCGCGGGG
AGACGGCGGTGGACAC

65 SEQ ID NO:15 orf6para

ATGCGTGCCCTCTTCGCCCCAGAGGGTTCCGCGTGCACCACGGTCACGCCGGGATCAGGGGGTCCACGCGGACCTCGCCGT
CATCGCCTCCGACGTTCCCGCGGCGGTGGCGCGGTGTTACCCGTTTCGCGGTTTCGCGCGCGGAGTGTGCTGCTCAGCC

GGGACGCGGTGCGCCGACGGGATCGCCCGGGGCGTGGTGGTGTCTGTCGGCAACGCCAACGCCGGGACGGGCCCCGCGGGG
TACGAGGACGCCGCGGAGGTGCGCCATCTGGTGGCGGGATCGTGCAGTGCAGAGAGGATGTGCTGATCGCCTCCAC
GGGACCCGTGCGCGAGCGGTATCCGATGTCCCGTGTCCGGGCCATCTGCGGGCGGTGCGCGGGCCCTTACCGGGTGCCG
ACTTCGACGGCGCGCGCGCGCGCCGTGCTGGGCAACGCGGGCGCCCGTCCACGATCCGGCGGGCGCGGTGCGGCGACGCG
5 ACCTGATCGGTGTGCGCAAGGGCCCGGTACGGGCCGCGCGAGCAGGACGACCGGTGACGCTGGCGTTCTTCTGCAC
GGACGCCAGGTGAGCCCCGTGCTCTCGACGACATCTTCGCCCGGTGCGGACCGCGCCTTCCACGGGTGGGCTTCG
GCGCCGACGCTCCACCGCGACACGGCGCGCTTCTCGCAACGGGCTCGCGGGCCGGGTGGACCTCGTCGCGTTTCGAA
CAGGTCTGGGCGCGCTGGCGCTGACCTGGTCAGGACGCTCGTCCGGGACAGCGGCTGCGGCGCGCCCTGGTCACGGT
10 GCGGGTACCGGGGCCACGACACCGAGCAGGCGCGGCGGTGGGCGGGCGGTGGTCGACGCGCGTCTGTAGGGCCG
CGGTGACGGCCCGGACCCGACTGGGCGCCGTGCGCGCGGTGGCGGGTGGACACGGGGACGAAGGCCCGCGCGGTCT
CCCGGGCGGATACGATCCGGGTGCGCGCGCGGAGGTCTTCCCCGCCCCCGGACCGGGCCCGCGGACCGCGTCA
CGGTATCCGACGGCGCGAGGTGACCGTCCATATCGACCTCGGTGTCCCGGGCGGGCGCGCCCGCGCGTTCACGGTCC
ACGGTTCGACCTCTGGCGGGTACCGCGCCTCGCGCGCGGCGCGGCGGTCTGA

SEQ ID NO:16 para cluster

CCATGGGAGCAGCATCGCAGTGGCCTCCCCGGCGCCATGCGCTAGTGGTAGTCCCCCTGCCGGGTGCCGACCGCCG
GGGCGGTCCCGGGTGGCGCGCCGATCTAGTGGTGTCTCCGACGGTGCTGCTGGGTGAGGGGCGAGTGTAGGCGGA
TGGTGGTTCCCGCGCGGGGCGGGCTGTGACGCGCAGTGGCCGCGAGTGCTCCACCCGTGGTGAGGCGGACGAG
20 CCGAGCCCCCGCAGGGGGCGCGCCACCGCGCGGTGCTGGCGCGGTGGTACGCGTGGAGCCGTCCGTCCCGGGTGGCCAC
ATGACGTCGCGCGAGTGGCACCGGAGTGCTTGGCGCGGTGGTACGGGCTCGGAGACGGCGTAGTACGCGCGGTCT
CGACCGGTTCGGGGTGGCGTTCCCCGGTCTGGATGTGAGCCGACCGGGATGGCGGAGCGCGGGCCAGGGCCTTGAGC
GCCGGGCGGAGTCCGCCCTCGGCGAGTACCGCGGGTGGATGCCCCGGGCGACCTCCCGAGTTCTGTCGACGGCGCGGC
CAGCCCGTGGTCACTCGTCGAGCTGCCGATCAGTCTGTCGGCGTCGAGCGGCACCGACAGTTGACGGTGGCACC
30 GCAGCGCCAGGGAGACAGGCGCTGTTGGGGCGCGTGTGAGTTCGAGTTCGATACGGCGCGCGGTGTGTCGGCGCG
GCGACGATCCGGGCGCGTACGCGGTGAGGCGCGCTGCTGCTTCCCGGTGGCGATGGCGGTGGCCACAGTTTGGTGAA
GCCGGCCAGCGGTCTCGGTGTCCGACGGCATCGGCTTGTCTTTCATCGACGCCACGCTGAGCGCGCCCCACAGTTGTC
CGTCGACGTTGATCGGCATGCACACCGTGGCGCGGAATCCCCACTCCTTGGCGACGACGAGGCGGGGCCGAGGACAG
GCCGCGTAGTCTGTCGATCCGCGCGGGCAGCCCCGACTCGAACACAGGGTGTGCACATTCCGGCGCGCGGGCGGTACCTG
40 GATACCGCGGGGAAATACCGGCGGTCTGGTCCAGGCGGGACATACAGGGCGGTTCGGTGGGCTCGTAACGGCGGA
GGACCGCAAGTCCGGCGAGAGGAGCTGTCCGGCTCGGCGCGGACCGCGCGGAACACTCCTTCGGCGGTGCCGCCGC
GCGACAGGGTCCGACGCGCGCAGCGCGCCTGCTCCTCGGCGCGCCCCCGCAGCTCCACAGTGCCTGGGTGTTTCGC
GATGGCGGTGGCCACGAGGTGGTGAACCGGCGAGCGGTCTCGGTGTGCGGCGCGAGCGGTTCGCGGTGACGAGGA
TCGCCATCATACGCCCCACAGCCGTCCCTCGACGTTGATCGGCACGCCGACCGAACCGAAGCGCGCGCCCTGGCG
AAGTCGGCGGGTGCCCCGAGCACTCGGCGCGCTGTCGATCCGGCGCGCGCCCGTCTCGGACACAGCGTCAACAC
35 GTTCCGGCGGTGCGGGTCCACCGGGTCCGATGGGGAGAGCGGGCGGTGACACTTCTGGACAGCGCGCGAGCGCGC
TCGCCATGGCTCCGGATCGAGCCTGATGATTCGGGTACATCGTTGCCGAGCAGTTCTCCGACTTCGGCGCGACCGTC
GCGAACATCTGTTCCGGTGGGTGGCCCTGGCCACAGGGTCCGACCCGTGCGAGTGCAGCGCGCTCCTCGACGATCTG
TTCGCACGACACGACCGTGCAGGCCCCCTACCGCCCGATGACGCCCCGATACCGGGTATACCGGCACATCAGCATG
ACGTCGCCCGTGAACGCCGTCAACGTGGCCCGCGGAGTCGGGAACACGCGTCCGGAATCAGCCCCGAGCGGGA
40 CAGCGGTTCCGGTTGGTGGGAAGGATGTTGGCGGTGGCGCGGATGCGGAAGCGGATCGTTCCAGTACTTCTGGGAAG
TGCGTCGCGGAGAGTCCGTCCGCTCCCCGAGTGGGCGCGACGACGCTGCGGGTCTCCACGGGGAGAGATCCGCGAA
CCGGCGAAGGAGCTGCCGTGTGCGACGCTTTCGATCCGAGAAGAGTTCCGCCGGTGTCCGACCCCGCGCGGCAACGTC
CCACCGCGCTCTGTCATCAGCGCGCTCGGCGCGCTGACCCAGCAGAGAAGATCGGATACGAGTGTACGAGCGGGA
45 TGAGGTTCTGTCAGCAGTCCCGGCTGCGCGCTGCGGCTCCGTACCGTCCCTGGGCTGTCTGGGCGTACGCGCG
ACGGCCGGAACCTGGAGCTGGGCCCTCCCGGTGAGCGGGCGGTTTTCGCCCTGCTGCTCATCAACGCGGGCAGTGTGGTG
CCGGTGCAGTCTGATCTTCCGTATCTGGGGCAACTACACCGGGCGCGGTACCGCGACGCTCCAGTCTATGTGTC
CCGGTGCAGAACTCCTGGCCGAGTGTGTCTCCCGACCGTTTCGACACCCGAACTGCTGCACCGCGCGGGCTTACA
50 CCTCGCGCTCGGACCGGAGCAGATCGACGCGAACGTTTTGAGCAGGCGATCAGGACAGGGCGCGGCTTCGCGCGAG
GAGCAGCACCAGGAGCGCGGGCGGTGCTTGCAGGCGCTGTCGCGGCGGGGCGGACACCGTACGAGGAGCTGAGCGC
GTACGACTTCGCCGTCCAGGAGGCCAATCGGCTGGAGCAGTCCGGCTGGGCGCGGTGGAGACATGGGCGCACTGTGTC
TGCGGTGGGGCGGGACGAGGAGGTGATGGACAGCTCAAGCCGAGGTGACGCGCAATCCGCTGCGGGAGCGGTGATC
GGGACGCTCATGACGGCGAGTACCGGCTGGGGTGCCAGGCGGACGCGCTCAGGACGTACGAGGCGACGCGCGGGCCCT
55 GGCCGAGGAGCTGGGGACCGATCCGGGCAAGGAGCTGGCGGCGGTGACGCGCGGATCCTGCGTCAAGACAACGGTCTGG
TCGAGGCGGTGACGCGCGCGGTGGCGGGGCGGGCGGGCGGTGACGGTGTGGTCCCGGCACAGCGG
CCCCCGCTCCGCTCCGGTCCGTTTCCGCGTCCGTTTCCGGCTCCGGCTCCGGCTCCGGTCCGCTCCTGCGTCCGTTT
CCACCTTCTTCCCGGCTCCGTTTCTGGCTCGGCGTCCGTTGCGCGTCCGTAGCCGCGCCCGTTTCCGGCCATGTCTCC
60 GGGCCCGGGTCCGCTTTCGGGTCCGTGGCGCTCCACCGGCGCGAGCCCTCCGGGGCGAGCCGTTCCAGGGGGCGCGCA
CCGCGAGTCCGCGTTCACACCTCGGGCGGGTGGCGCTTCTGTCGCGCGCGGCGGACGAGTTCGCGCGTCTGTCGAGT
TCCGAGTTGGAGCGTCCGTTCCGGACAGTGTGCGCACCGTCTGGGCGTCTGTTCCGAGAGTGAGGACCGGGCCGACTA
CTGGCGTGGACGACCGTGTGCGGCATCTGTACGCGATGTGGCCGGAACGTATGACGAGTTCCTCCGGTTGGTGTGCGC
GCGCACTCGCGAACTGCTTCCCGAGGTGGGCGCGGAGCCACAGGGGCGCACTCCCCGACGGGGCGAGGAGAACAGC
65 GGCAACGGGGACGGTGGGGCGAGGGGACAGACCCCGGCGCACCCCTCAGCTCGCGCCCGCTTCCGCCCCCGCG
CTCCAGAGAGGCTCGTTTACCTGACGACGCGCTGTGCGCGCGCTTCTGCGCACGGTCCGCGACCCCGTGGTGATCA
TGCTGGAGGACATGGAGCGGGCGAGCCCCCTCGTCTGCGCTCCTGGTGGAGCAACTGCGCACCGTCCCG
CTGCTGCTCGTGGTACCAACGCGCACCTTCCGGCTCGCGCACGACCGGAGTGCAGCGGGCGCGCGGTGATCCTCCA

GTCGACCGGCGCGCCGGGTCTGCTGAACGCCCTGGACGCACGGGCCACCGGGAACTCGCCGGAGGGATGCTGGGCA
 AGGCCCCGGACACCTCTCTCGTACGGGGCCCTGCACGAGCGCTCCGCCGGGAACCCGTACTTCTCGTCCAGCTCTCCGC
 TCGCTCCGGCAGGGGCTCGCCGCCGCTGGGAGACGGAGATCCCGACGAGCTGGCCGGGGTCTGTCTGCAACGGCTGTG
 5 GAGCGTGCCGCCCGCGTGCGCCGGGTGCTCGACATCTGCGCGGTCTGGAGCGCAGTTGCGAACCGCTGTGATCGAGA
 CCGTGCTGCGCCATGAGGGAATCCCGCTGGAGAACGTCCGTACGGCGGTCCGCGGCGGTCTGTCTGGAGGAAGACCCCGAC
 GACCCCGGGCGGTGAGGTTCTGTGCATCCGCTGGTCCGGAGGCGGTCTGGGACGACCTGGAGAACACCCGTCCGCCCGT
 GTCCCGTTCTCCGCGCTCGGGGCGCTGGCCACGGTCTGAGTCCCGGGCCCGGGGTCTCTGGCGCGGGCGGGCGCTTGC
 GCGCTCCCGACGCGGGGCTTGATCCCCCGGGGACGCCGACGCGCAGCCGGGTGCAAGGGGCGGTGCCGACACTGGGCG
 10 GGCGGCGGCCGTGGCCGGTGCGCCCCCCACGGCCACCAGGAGCCCCCATTTGGACACGTACGACGCGGATACGTACC
 CGCGGTCCGGACACCCAGCCGAGCGCGTCCCGACGCACCTCCCCACGCGGTCCCGGGACCCGTCCCGGACCCGTTC
 GAGCCGCGCCCGGACCCCGGGCGCGAGGCGCGTGGTGTCTCGCGGCGGACCGCGCCCATATGTTCCACCCGGTCTTGCC
 CCGGGGCGCGAGGACCGCACCGTTCTGGTCTCGGCCGCGGTGCACCGTACGGGACACCGAAGGGCGCACCTATCTCG
 ACGCTCTGTGCTGCTCGGACTGACCCAGATCGGCCATGGACGTGAGGAGATCGCGCAGGCCGCGCCGAGCAGATGCGG
 15 ACACTCGGTCACTTCCACACCTGGGGACCATCAGCAACGACAAGGCCATCCGACTGGCCGCGCGCTCACCGACCTGGC
 GCCCCAGGGTCTCAGCGCGTCTACTTACCAGCGGGCGGGCGAGGGCGTTCGAGATCGCCCTGCGCATGGCCCGTTACT
 TCCACCACCGCACCGGACCGCGAGCGCACCTGGATCTTGTCTCGCGCCGACCGCTACACCGCATCGGCTACGGCAGC
 GGTACGGTGTGCGGCTCGCCCGCTACCAGGACGGGTTCGGCCCGGTGCTGCCCCATGTGCACCACTACGCGCGCCGA
 CCCGTACACGCGGAGCTGTACGACGCGGAGGAGTACGAGTACTGCGCTGCGCGAACTCGCCCGCACCATCGACGAGA
 TCGGCCCGGGCGGATCGCCGCGATGATCGGGGAGCCGGTATGGGCGCGGGCGCGCGTCTCGCCGCGCGGCGGACTAC
 20 TGGCCGCGCGTTCGCCGCGTGTCTGCGTCCCACGGCATCTGTCTGATCCTGGACGAGGTCTGTCACCGCTTCGCCCGCAC
 GGGGACCTGGTTCGCGGCCGAGCACTTCGGGGTGACCCCCGATCTGTCTGGTGACCGGAAGGGCATCACCTCCGGGTATG
 TCCCGCACGGGGCGGTGCTCTGACCGAGGAGTTCGCGACGCGGTGAACGGGGAGACGGGGTTCGCGATCGGCTTCACC
 TATACCGGTCACCCCACGGCGTGCGCCGTGCGCTCGCCAATCTCGACATCATCGAACGGGAAGGGTGTCTGGAGAACGC
 GGTGAAGGTGGGCGACCACTCGCCGGGCGGTGGCGGCCCTGCGCGGGTTCGCCCGGTGGGGGACGCTCCGGCACTGG
 25 GCATGATGCTCGCCGTGAGCTGGTGTGCGACAAGACGGCCCGCACCCCGCTGCGCGGGTTCGCCCGGTGGGGGACGCTCCGGCACTGG
 GCGCTGCGCGAGGACGCGGGCGTCTGCTCGGGGCGCACCGCGCTCCCTGGTCTCAATCCGGCGCTGTGATGGACCG
 GGCCACGGCGGACGAGGTGGCGGACGGGCTGGACTCGGTGTGCGCGGGTGGCACCCGACGGGCGGATCGGCGCGGCC
 CCCGGCGGGGTGACGAGACCGCGGGCGCCACCCGCGGGGGCGCGGGTTCGGCACAGCGGCCGACCCGGCGCTTCCC
 CGTTTCCCGCGCGCTTTTCCGTGCCCCGCGCGCTTCCCGTGGCCCTGCCCCCTGCCCCCTGCTCGGGCGCTCTCTCCCTC
 30 GCTGTGGCGCGTTCCTCGTTCCAGCGCGTGTGCGAGCGCGCCCAAGCGCCCCGTGCCACGTTGGGAGACCGCGCCG
 CGGGGCGCGGGAGCCCGCAAGCCGAAGGAGTCCCGTCCGATGCGTGCTTCTTCGCCAGAGGGTTCCGCGTGCACC
 ACGGTCAACCGCGGATCAGGGGGTCCACGCGGACCTCGCCGTATCGCCTCCGACGTTCCCGCGCGGTTCGGCGCGGTG
 TTCACCCGTTCCGCGTTCCCGCGCGGAGTGTGCTGCTCAGCCGGGACGCGGTCCCGACGCGGATCGCCCGGGGCGTGGT
 GGTGCTGTCCGGCAACGCCAACGCCGGGACGGGCGCGGGGTACGAGGACGCGCGGAGGTGCGCCATCTGTTGGCGG
 35 GGATCGTCACTGCGACGAGAGGGATGTGTGATCGCCTCCACGGGACCCGTTCGGCGAGCGGTTCGCGATGTCCTCGGTG
 CGGGCCCATCTGCGGGCGGTGCGCGGGCCCTTACCGGTGCGGCTTCGACGCGCGGCGCGGCGGTGCTGGGACCCGC
 GGGCGCCCGTCCACGATCCGCGCGGCGCGGTGCGCGGACGCGACGCTGATCGGTGTGCGCAAGGGCCCGGTACGGGCC
 CGGCGGAGCAGGACGACCGGTTCGACGCTGGCGTTCTTCTGACGAGCGCCAGGTGAGCCCCGTCTCTCGACGACATC
 TTCCGCCGGGTTCGCGGACCGCGCTTCCACGGGCTGGGCTTCGGCGCGGACGCTTCCACCGGCGACACGGCGGCGGTCT
 40 CGCCAAACGGGCTCGCGGGCCGGTGGACCTCGTCGCGTTCGAACAGGTCTTGGGCGCGCTGGACCTGGAGTTCGAGG
 ACGTCTCCGGGACAGCGGCTGCGGCGCGCTGGTTCAGGTGCGGGTTCACGGGCGGCTTCACCGGGCCACGACACCGAGCAGGCGGG
 CGCGTGGCGCGGGCGGTGGTTCGACGCGCGCTGAGGGCGCGGTGACGCGCCGGCACCCGACTGGGCGCGGTTCGC
 CGCCGTGGCGGGTGGACACGGGGACGAAGGCCCGCGGTCTCCCGGGCGGATCACGATCCGGGTTCGGCGCGCGGGAGG
 45 TCTTCCCCCGCCCCCGGACCGGGCCCGCCCGGACGCGCTCACCGCGTATCCGACGCGGGCGAGGTGACCGTCCATATC
 GACCTCGGTGTCCCGGGCCGGGCGCCCGGCGGCTCACGGTCCAGGCTGCGACCTCTGGCGGGGTACCGCGCTCGG
 CGCCGGCGGGCGGTCTGAACGGGCGCTCCCGGGCGGACGCGGACCGCGAGGGCGCGGGAGCGCAGGGAACACGGGAGCG
 GGCCCGGTGCTGCGCCACCGGGCCCGCTCCCGTCTCGTTCGTCGCGTCTCCCGGCGCGCTACCCCCACCGTGGC
 CGGCGAAGTCCACGCGCTCTCGGCGTCCACCGCTCCACCGGTTCTCGGCGTCTCGGCGTCTCGGCGCGCGCCCC
 50 GGTGGCAGGGGAGAGTCCACCGGTGCGGACGCGGGCGACGTGGTGGCGCGGGCGTACTGGTAGAGCAGTTCCGGCCCCGAT
 CTCCGCGCCAGCAGGGAGGTGATCCCCGACGGGTCTGACCGGGGACACCTCGACACGTGCAAGCCGACGGGCGCTGA
 GCTGCCCCGACACGTGAGCAGGGTACGACCTCGCGCGAGGACAGCCCGCGGGGCGGTGTGCGCGGTGCGCGGATCTGCTCGGCGAT
 TACGCCGGGTGAGCAGCTGATGTGACGAGAGACGTACAGCGGAGGCGCGGACGCGGTGCGCGGATCTGCTCGGCGAT
 55 GCCGCGCGGTGAGCGCGGGTGAAGTTCGGCGCGGTGACGATGCTGACGCGGTGCCCGCGCGGTAGTCCAGGGAGTCCG
 GCCGCGGATTGTGGCCGCGGATGCCGACCTGGACAGGCGCTCCGGGTCCACAGGCCCTCTTCGATGGCCAGCGGAAG
 GGGGTGCCGTGGTGTAGGTGCCGCGGTAGACGGGTGGGTGGTGTGCTGTGCGCGTCCAGGTGACGAGCGGCGACCCG
 GCCGTGGCGGGCGTGCACGGCGCGCAGGGCGGCCAGGAGAGCGAGTGGTCCCGCCAGCATCAGGAACCGGCTCGTTGC
 GTTCCAGGAGCCGGTCAAGGCGACCGTTCGCGGTGTCATCGCCAGGTCCATCGAGAAGGGGTGAGGTGATGTGCCCC
 60 CCGTCGACACGTGATCCGTCGAAGACCCCTGGGCCCGGTGATGCCGACGCGGTGGATCAGGCTGGACTCGTGGCG
 GATGGCGCGCGCGGAACCGCGCGCGGGCGGTAGCTGGTGCCTCCGTGATCGGGGCGCGGACGACACCACTCAT
 GGCGGATCGGGTGGGCGGTGGCGCAGCCGATGAAGGTGCGCGGTTGGGCGTAGCGGGGAGACGGCGGTGGACACC
 CTGGCGGTTCCCCGCGCACCCGGCCCTGCTCCGTTCCCGTACCGACGCGCCCGTGGATCAGGCTGGACTCGTGGCG
 65 GCGACCCCGTTCCCCGAACGGGCTCCGTTCCCGTGGAAATCCCGTTCCCGCGCCCGCGCGCGGTTCGGGCGCGCGG
 TGCCCCCTCCCTCGAGACCGCTCCTGCGGCTTCTGCGGCGGTGGCGCTCTGCGGGCGGTGCCCGCGCCACGCGCGT
 GCACCGTCCGCGCGCGCGCGGTGCGGTTGCCGCGCGGTGCGGTTCTGGCCACCGGTGCCGTTCTGGCCGCTCATACG
 ACCACCGGCGCTGGAGCCTGAGCCTGCGCACCGGTGACGAGCGCGGACCGGTCTCGCGAAGTCCAGCTCCTCCGG
 CGGACCGTGTGATGACACCGGTGATACAGGCGCGGTGCCATGGCGCGCTTGAACGGCGTCACTCTCGCGCGCGGA
 TCCCTTCGGCGAGGAGTCCGTTCCACGCGTGTGGTGGCGGACCCCTCGTGGATGCCAGCTTGGGGCGGGCCAGG
 GTCTCGGCGGGCAGCAGGCCGAGAGGGCTGCCGAACACCCACTTGTGCGTGCCCCGCGCGGTTTGAACCGGGTTC

GAGGGAGACCAGCGCTCCAGGACCGCGCGGTCCAGTACGGGTGGGTGGTCCACTTCCCGCGATGCCCGAGGACGG
 GGGACATCTCGTTGAGGCCGTGCAAGCCCGCCATGTGCGCCGCGATCTCGTCTGAGGGACAGAGCGAGGCCGTGCGC
 CGGTGCATACCGCCGAGCGGGATGTGCGCGCCGTACCCGGTGAGGATGCGGAGCGGCCCGGTGTGAGCCGCCGTAGAG
 5 GGCAGCAGCGCGCAGGTAATCCAGGACCGTGGGGTGGTGATCTCCGCGCGCGGACCGCCAGGGCAGTTCCCTGA
 CGAGTTCCGCCGAGTGGAGCCGATCTCGTGTGCGCGGTGCCAGGTGGACGGCGACCGAGCGGGCCGCTGCAACTCG
 TCGGACACCTCGGTGCCCATCGACACGGACCGTGTCCCGGTGCCAGGGCCGCCGTGTGGGCGCGACTCCCCCGGAGTC
 GATGCCGCCCGACAGGACGACGGTGGGGGCCGCTCCCCGCCGCGCAGCCGGGTGCGGACCGCCGTGGCGAGGCGTTGCG
 CGACCAGGTCCACCGCTCCCGTTCCGCCGGGCGAGCGCCGGGAGAGCGGGGGTGTCCAGGTGCGGACCGCCCTGGCGGTG
 10 ATGTGCGAGCCGCCACTCCGTGCGAGCAGGAGGGCGGTCCCGCGCGGGACCCGGCAGACGCCCGCCCGCCCGCGCGGT
 GTGGGTGCCGACAGGCCAGCGGCCCGGCCGCTCGTGCGCCAGGGTCTTCGCTCGGTGGCGGCGCTCAGCCCCGTCA
 CGTCGGCGCGCAGCCACAGCGGTACCGAACCCGCGTGGTTCGTTGGCCGCGACGGTTCGCGCGGTGGAGGCGCTCGGTGAGC
 AGTGCGGCGAACCGTCCGTTTCCAGAGCCGGAAGGCCCGGGGCCAGCGCCGCGCAGGCGGCCAGCAGCAGTTCCGCGTC
 GCCGAGGGCGGCGAGGAGCCGCCGAGCGTCCGGTCAGCTCGGCGCGGTGTACAGCTCGCCCGCCAGGAGCAGCCGGA
 CCTGGCCGTGCGGCGACAGGACGGGCGGACGGCCAGGGTCACGGCGGTTCGCTCCAGAGCGGGTACGCGGTGCCGTG
 15 TGCACGGGGACATGGGTCCCGCGGACGGCGAAGCGGGGTGCGTCCCGGGTTCGGAGTGACCGCCGGGGCCGCCCGCGG
 CGGCCCCCTCGGTGCCGATCGGCACCCGGAATCCGTACACGAGGTCCGGGCGGGCGCATGGTGAATCGTCTCCAGGTGG
 TCAGATGGCCAGGGCGCGAAACCGCCGACTGGAGTCGTAGGCCACCGGTACCTCGATCAGGAACGGGCGGCCGAGTC
 CGGCGCCCTTGGTGAGGGCGGCGAGCAGCGAGGTGCGGTGCGTGGCGCGGACGGCCTCGCAGCCGTGGCCTCGGCGAGC
 TGGACGAAGTCGACGCTTCCGAAGCCGACGGCGGGGGCGTGGGAGCGTGGTGTCCGAGGTTCTGGTACAGCTCGATCAG
 20 GCCGTTGCGGTGCTTGTGACGACGACCATGACGATCGGCAGGCCCAGGCGCACGGCCGTCTCGATGTCCGCGCTGTTGG
 AGTGAAGCCCGCTGCCCGCGATGAGGAAGACGGGTCCCGGGCCGGGCGATCTGGGCGGCCATGGCGGCGGGCAGT
 CCGTAGCCGGAAGCTGGAGCAGCCCGCGGAGGTGAGGAATCCGTACCGCTGGTTCGGACTTGGCGAAGAGCACGCCGTAGTG
 GCGGAAGAAGCCGATGTGCTGACGAAGGTGCCGTTGTGAGGACGGAGTTCATGCACTCGATCACCTGGTGGACCCGCA
 TGCCGTCTCGTACTCGGTGGGGTGGCGAGGAATTCGGCGACGCGGGCGCGCAGGGCGCTGAGGTGCTGCCGGGTCTTG
 25 GGGGCGAGGCCCGAGGTGCGTCTCGTTCGAGCGCGGTGACGAATTCGGCGACGTTGGTGACGATGTTCGATGTCCGCGCGGA
 CAGTCCGGGATCGGTTGACCTCGGGGGCGACCCGACCGTGGTCTTGGCCCGGCCCGCGTCCACATGGAGGGGCGCA
 GGTCTCGGCGTAGTCGTAGCCGATCGCCAGGAGGTTCGGCGGGGCCGGAAGATCTCGTCGAGGGCCGGGTGGCCGAGA
 ATGCCGTCCATGTAGCCGCTGATGGCGCGTAGTTGAGCGGGTGGTTCGTCGGCAGGACGCCCTTGGCGGTGTAGGTGGT
 GACGACGGGGATGTTTCCGCGCTCGGCGAGGGCGCGCAGGGCGTTCGAGGCCCGGGCGCGGATGACGGCGCTACCGACGA
 30 CGAGGAGGGGGTTCGCGCCTCGCGCACAGCTCAGCGCCCTCGTTCGAGGCGGGCGCGCCAGTCCGCGTCCAGGGCGTGG
 GTGGCGGTGGCCCGGACAGGGGGCGTGGTGGGGTGGCGTTCAGTTCGGCGCCGAGGAGGTGACCCGCGAGGTGAT
 GAAGTGGGACCCACGGGCTCGATCCGGCTGTTGAGGACCGCGCTGTGACGAGGTTCGATGTCCTCGCCCGCTTCGA
 GCTGGACGCTGAACCTGGTTCAGCGGGCCCATCACGGCGGTGCTGTCCAGGCACTGGTGGGTGACGTTGGGGTAGCAGTCG
 35 TACGACTCGGACTGCGCGGCCAGCGCATGACCGAGCTGCGGTCCAGGGCGGAGGTGGCGACGCCGTGGCCAGGTTGGT
 CATGCCGGGGCCAGGGTCGCGAAGCACGCCCTGGGGGCGGTTCGTTGATCCGGGCGAGGACGTCGCCATCACCCCGCGG
 TGAACCTCGTCCGGGTTCAGGACGAAGTCGAGTCTTTCGACCTCGTTCGAAGAGAATGGCGGACGCCCTCCCGGCCGACGACG
 CGAATACATAGTTCGACACCGTACTGGTGAAGACGTTCCAGCATGGCTTTCGCGGTTCGTTGGTGGCCATGGAGATCTCCTT
 CGCATCGGACGGGCGCGGGATGGCGCCCCGGAACCGGGCACCGGGCGGTGGCGACCGGGTGGCGCACACCGTGGGTG
 40 GTGGCGTTGCCACTGTGCGGATCGCCTCTTGGCGGCGGTTCGACGCCCGGCTTGGACAGAATGGGCAAGGCGCGTTCAAG
 GCATGGCGTCCATCGTCTCGTGGCGCTTTTCGTGAAATCCGTCCGGCGCCGACGGTCTCCATCCGATTCGTCCTCCCTTC
 CGTCCACCGATCCGAGGAGAATCCATGGATGTCTGGCGCGGTGGAGCGCAAGCCAGCCTGAATCTTTTCCCATCGA
 GACCCGGCTGTGCGCGCCAGTGGCGCGTGGCCAGGACCGCGTCAACCGCTATCCGTACTCCGAGACCCCGGTGG
 45 CCGTCTACGGCGATGTACGGGGTGGCCGAGGTGTACGCGTACTGCGAGGACCTGGCCAAGCGCTTCTTGGGGCGCGC
 CACGCCGGTGTGCACTTCTGTCCGGTCTGCACACCATGCACACCGTGTGACCGCCCTGACCCCGCCCGGCGGGCGCGT
 CTTGGTCTCGCGCCGAGGACGGCGGCCACTACGCCACGGTGACGATCTGCGGGGGTTCGGCTACGAGGTGAGTTCT
 TACCTTCGACCGCCGACACCTGGAGATCGACT

SEQ ID NO:17 cvm cluster

GGTAACCGGCATCCGACCCAGGCCCGGGCGCAGGACCCGGAGGCGAGGACCCGGCACACCC
 50 CGGCCGGGCGGCCCGGTCCCGGCGGTGCTGTCGGCGGACCCGCAATCGGCAGCCGCC
 CAGGCCCGGACAGGAGCCCGCTCAAGGCACCGGCCCTGCGCACCCGCTGAGGCGGCGAG
 GTTCTTGACAGCCGGCATCCGCCAGTCGGCGCGGGGCGAGCCGCCCCAGGCGCCCGGCCG
 GCACACCCGTGCGAGCGCCCGGTCCCGGCGGTGCGTCCCCCGGAGGCGGCGACCGGCGAG
 55 CCGGACACCGGCCCGCTCGGGCGCGGCCAGGGCACAGGCCCTGGGCAACCGCTCGGAC
 GCCCCTTCGGACAGCAGGCCCGTGGGAAGCCGCGGTTCAGGCCCGCAGGCGACCCAGGT
 CGGCGGGCGGATCAGGTGTTGGCGGGGACTCGTCCGGGAAGATCTTTGTGACGACGGTC
 CCGTCTCGGTGAGATAGCCGTGACGATCCCGGGGTGCTGTGCGGCGCTCGAAGTCG
 CCGGGGGGTGCGAGGGCGATCACGCCGCCCTGCCCGCGAGCCGGGGCAGGCGCTTGACG
 60 ATCACTCTGAAGCGGCGGACGCCAGCCGAGCCCTTGAACCTCGATCAGATGGGAGAGG
 GTCGAGGTCCCGCGCCCCGGATGAACACCTACCGGCGCGGTGGCGCTCGCGGCGAGC
 GTCCGTTGTGCGCGTAGGTCCCGGCCCGATCAGCGGGAGTCGCGGATCCGCGCGGGG
 AGCTTTGGTGAGCCCGCGGTGGAGGTGGCCCGCGGAGATCCCGCGCCCGTTCGAGG
 GCCACCGCGCCACCGTCCCGTCTGACTGCGCGTGGGCCAGTGCTCCGGGGCCCTCCGG
 65 GCGGCGGGATCGCCCGCTCGGTCTCTTCGCGCGCAGCAGCGGTCCAGCGGGCCTGG
 GTCCAGTAGTAGTCTGGGTGACGTTGCGCAGCCCGTGGCGGGCGCCGAAGTCGTGCGG
 CCTCGCCGGAGAGGAGGACGTGCTTCGACTTCTCCAGCACCAGCCGGGCGCCCTCGACC
 GGGTTGCGCAGGGAGGTGACCCGGCGACCGCTCCCGCTTCAGATCGGAGCCCGCATC

ACGGAGGCGTCCAGCTCATGCCCGCGTCCGGCGGTGAAGACGGCGCCCTTGCCCGCGTTG
 AACAGCGGGTTGTCTCCAGTTTCGCGGACGGCGGCCTCGACCGCGTCCAGGCTGTCCCCG
 CCGCGCGGAGCACCCGCTGTCCGGCGGAGCGCTGCGGCGAGCCGTCCTCCGGTACGCC
 TTCTCCCGTTCCGGGCGCGTCTCTCCCGTCCAGGGCGGCTCCGGCCCCGCGTGGACG
 5 GCGATGACCACGTACAGGGCGTCCGGCGGGGCTTCCCCGGCGCGTCCCCCGTTCTTTC
 TTCTCCTCCCGCGCTGTCTCTCTGCTTCTGTTGCGTCTGTGGGGCGCGCGGTGGGT
 CCATGGCCGCCGAGGCCCGGGTACGACGATGAGCGTGGTCTGACACCGCGCGGGCG
 AGCAGGGAGGACGCCAGCCAGGCGGTGGCGGGCGGTGGGGCATCGGGCACTCCTCGGA
 CGGGGGTGAGAGACGCTCCGGCCGACTGTACTGACATGCCCATGCCCTCTAGTGCCCC
 10 GGAGCCGCTTCCGCCCTCCCCGCCGCCCGGGCGCGCCCGCGCGCTCAGTCCAGG
 GCCAGGTCTCCGGGGCGGAGCGGGCGAGTCCGGCGAGTGTGCGGAGCGCCCGGTTCAGT
 TCGTCCGCCGACGGCGACGCCAGGCCAGCCGACCGCGTGCCTGTACGCCCTGCCCG
 CGCGAGAAGCGGGCGGGCGGTACCCCGATCCCGTCCCGCGCGGGCGGGCGACGAAG
 GTGTCCGCGCGCCAGGGGCGGGGCGAGCACCCACGACGTGGTACGAGCGGGGTTCGCC
 15 GACACGGCGAAGCGTTCGAGCGCGCGCGGGCGATCTCTGCCGTACGCCCGCTCCCGC
 CGCTTGGCGCGTACGAGCGCGTTCGCGTCCGTTCGCGAGCGGACCGCGCCTCC
 AGCGCGAACCGCGCGGGGCGAGACCGCGGAGCGCAGCGCGCGCGACCGCTCCGTTCG
 AGCCCCGGGGCACACCGCGAACCCAGGGTCAGCCCCGGGGCGAGCCGTTGGAGAGG
 CTGTCCGACGAGCACCGTCCGCCCGGGGCGACCGCGCGAGCGGAGCCGTGCCCTCCCGC
 20 AGGAAGCCCCAGACGGCGTCTTCGACCGCGGAAGGTCCAGCCGTCCAGGACCGGGCG
 AGCTGGGCGAGACGCCCGTCCGACAGGGTAGGGAGAGCGGTTGTGAGGGTGGGCTGG
 ACATAGACCGCCCGAGCGGAGCGCTCCGTTGGCCTCGTCCAGCGCCTCCGGAATCAC
 CCGTCCCGCTCCATGGCGAGGGGACGAGCGTGTGCGGAGCGGGCGCGATCGCCTTG
 ACCACGGGTAGGTACGCTCTCGACCCCGAGTCCGCGCCCGCGCGCACGAGCGCGCG
 25 AGCACGGCGGAGGTGCTTCCGACCGTTGCCCGCGAACAGCACCCCGCGGGGGTCCCGC
 CGCCAGCGCCCCGGGCGAGCAGCCCGCGCGCGCCTCGCGCGCTCGGGGGTCCCGCG
 GCACCGGCCCGCGGAGCAGGACTCCAGGACATCGGGCCGCGAGCGCCCGCGAGCCCG
 GTGGCCAGCAGCGCGCCTGTCTGGGGACGACGGGGTGGTTTCAGTCCAGGTTCGATCCGG
 CTTCCGGCGGGTTCGGAGAGCGCGGGCCGACGCCCGCGCGCGCGGACATAGGTG
 30 CCGCGCCCCACCTCGCCGACGGTTCGACCCCTTCGCGGGCAGCTCCCGGTAGACCCGGCG
 GCGGTGGAGTTCGGCGATGCCGACCCCGGGCGAACTCCCGTTCGCGCGGAAGCCGTTC
 CCGGGCGCGAGCCCGCCCGTCTCTGATCTCTCGCGGACCGCTCGGCCACCTGCCGTAG
 TCCTTCATCTCCCGTACCTCCCTGTCCGGTGGACCGCTTCCCGCCCCGCCCCCGGACC
 GTGAAACGGAACACCCCGTTCGGAGCTCGAGCTCCCGTCCGGAAGCTCCCGTCCGG
 35 AAGCTCCCGTTCCAGAAATGACCCGAGAGCAATATTCCCTATTGACCCGATCAAAACAC
 CGATCTACGCTCGGAATTGCCCTCACACAGACCGTTCGACGATCTGCCGACACCGGTACT
 GACGCCCCGTTCGGACCGCACCCGCGCGGAGCCGTGCCCCGCCCCGCGCTTCGCGCACA
 GGAGAGAGAAGGAGATGGTGGAGACGCGACTCGCCGGTGTGGTATGGTTCGCCCTCG
 GAATGGTCTTACCCCGGGACCGAACATGATCTATCTCGTCTCCCGACGATCACCCAGG
 40 GCCGACGTTCGGGGATCATCTCGTGGGCGGTGTGGCCCTCGGTTTTCTGGTCTATCTGC
 TCGCCGCGAATCTCGGCCTGTCTGGTATCTTCGTTCGCCGTGCGGAGTGTATGTTCGCG
 TCAAATGGCCGGTTCGGGCTATCTGGCATATCTCGCTGGAACGCCCTTCGCGCCCGGTG
 GCGTGAATGTGTTCTCCCGGAGGAGTTCCGACGACTCCCGGAGCAGGCTGTTCACCA
 TGGGGTGTGACGAACATCTCAACCCCAAGATCGCGTTCATGTATCTCGACTCATCC
 45 CGAGTTTCGTTCGACCCGAAACCGCGAGCGTCTCTGTTCCAGGGGCTGATTCTCGGCGGT
 TCCAGATCGCGGTGAGCGTTCGCGTCAATCTCGCGATCGTGTGGCGGCCGAGCCATCG
 CCGCCTTTCTCGGCCGCCACCCCTTCTGGCTCAGGGTTCAGCGCCGCGTGTGGGCGCG
 CGCTCGGTACGTCGCGGTCTCCCTGGCCCTCGACACCTCCGCCCGCGCGCACCCGTCT
 CCTGAGGCCCGGACCGGGAGCCGACGCGAAGGACCCCTGGGCAACGTTTCGGAGAGC
 50 TTATCCGTTACCCCATGATATCCGATATAAGTGCATTGGCCACTTACCCATGCATGGAAC
 AGGCCAACCTGACCAAAAAATGAGCCCTCCCCACCCGGAATAGATGCTTCCAGTGTGAA
 GAAATTTATAGCGGGAGCGTCTGCCGAACAGGACGGCCATACGCCGCAAGGCGAAGC
 GACATCGCCGCCCGCCCGGGTCCAGAAATTCGGAGGACACATCGGACGACCGTCTCCGC
 ATCGGCGTCAACTCCCGATTACAGAGAATATTGAGTACGTATCAACGGGCTTGATCTA
 55 CTCAGCCTCCATTGTTCTCTCCAGTCCGGATGTGCAATGAAGTACGACATAACCCACCA
 TCCGGCCTTCGGTTTCGACCTCTCGGCCCGTTGACCGTACCGCGCGGAGCAACCCGTG
 GACCTGGGCGCGCCACGGCAGCGCGCCTGTCTGCCCTGTCTCATCGATGTTCGGCAAC
 GTGGTCCCGCTGCCGTCATGACCGCGTTCGATCTGGGGGGCGACCCACCGTCCCGGGT
 60 CGGGGGACGCTCCAGGCTTATGTGTCCCGACTGCGGAAACTCCTGCACCGCCATGACCGT
 TCCCTTCGCTTGTCCACAGCTCCAGGGTATCTCTCGAAGTGGATTTCGGCGAAGGTG
 GACGCGGTGGTTTTTCGAGACAGTGTTCAGGGAGTTCGGGGAATTGAGCAGGGCCCGAAC
 CCCGAGGCCACCCGGGCGTGGCCTGGTCCGCCCTGGAGATGTGAAGGGCACACCCATG
 GGCGAGCTGCATGATTATGAATTTGTGGCGCGGAGGCGGACCGCTGGAAGGAATCCGG
 65 TTACGCGCGTGGAGACCTGGTCCCAGGCGTGTCTCGATCTCCAGCACTATGAAGAGGT
 GCATTTACGTCGGCGAGGATCCACCGCAATCCGGAACGGAACGGCTGGGCGGTCTC
 TTCATGCGGGGCGAGTATCATTCGGACGGTTCGCGGAAGCCCTGTGACGTATGAACGT
 ATGCGTACCGCGGTGGCGGAGAATCTGGGGCGGATATCAGTCCGAGCTCCAGGAACTC
 CATGGAAGATTCTGCGCCAGAACTCACGAGACACCCCGCGCGATCGACGGCTCC

CTCACACGGGCGGCGGGCCCGCACGGGCCCCCGCCCTGGCCGAAACCGGCACCCCGCC
 GCACCCGCGGACATGGCCGAAACACGGTGGCGGAGGAAAGCGCCGCGCCCCCGCCCCG
 GCGGCGCCCGGACCCCGCCCCCATGCCGTCCCCCGTACCGCTCCCCCATCCGTCAGGG
 5 GCGCTCCCGCCGGTACCCCGGTGCCCTCCCCCGGTCCCCCGTCCGGCCCTCCGTTACGCG
 GCACCCGCGGAGACCGAGGACCCGGAACCGGCGCCGCCCCCTCCCCCTCCGCGGGCGGC
 CGACTCATCGGCCCGCGCGCCGAAC'TGCGCAGGCTGCGGCTGCTGCTGACGAAGACCCGC
 GCGGGCCACGGCCATGTCTGCTGGTCTGCGGCGAACAGGGCATCGGGAAGACCCGGCTC
 CTGGAGCACACCGAGCACACCCTGGCCGCGGGCGCGTTCGGGTGGTCCGTTCCGACTGC
 10 GTCGCCACCCTCCCGGCACCGGGCTACTGGCCCTGGGAGCACCTCGTACGCCAGCTCGAC
 CCGGACAGCGGCCCTCGGTGACGACGGCGACGCCGACCCCGTCCGCCAGGCCGAGTGGCTG
 CCGGAACACCACCTCACCCACCAGATGCGGATCTGCGCGACGGTGCTCGCCGCGCGCGG
 CGGACCCCGCTCCTGTTGATCCTGGAGGATCTGCACCTCGCCACGCGCCGGTCTCTGGAT
 GTGTCCAGCTCCTGGTCAAACAGATCGGCCAGGCCCGTTCATGGTCTCGCCACCCTG
 CGCGAGCACGATCTCGCCCGGACCCCGCCGTCCGCGGGGCGGTGGGCGGCATCCTCCAG
 15 CGCGGCAACACCGGCACCCCTCCGGCTGGACGGGCTACCGAGGAGCAGAGCCGGGAGCTG
 ATCGTCTCGGTGCGGGGGCCCCGTTCCGCGCCCATGACGCCAACGGCTCCAGCGCGCC
 TCGGGCGGCAACCGTTTCTGCTGCTCAGCATGGTACAGGGGAGGACGGCACCCAGGAG
 TGGGCACGGCCGTGCGTCCCGTTCGAGGTGCGCGAGGTGCTGCACGAGCGGCTGAGCGAA
 TGCTCCCCGTCCACCAGGACGTGCTCAGCTCTGCGCCGTGCTCGGCATGAGCGTGCGC
 20 CGACCGCTGCTCACCAGATCATGTCCACGCTCGACATCCCGCACACCGCGCTCGACGAC
 CGCTCGGCACGGGGCTGCTGCCACGACCGGAACACCGAGCGAATGGTCCAATTTCGCC
 CATGGGCTGACCCGGGACTTCTGCTCGACGACACCCCGCGGTACCCCGCGCCCGTGG
 CACCACCGGGTCCGCGCCACCCTCGCCCTGCGCTTCCAGCAGGGCGACACCACGCCGAG
 ATCCGCCGCCACTGTCTGGCCGCGGCCCGTCTGCTCGCGCCCCGCGCGGGGTGCGCCCC
 25 CTGCTGGCGCTGGCCGACCGGGAGCAGTCCCGCTTCTCCACGCGGAGGCGCTGCGCTGG
 CTGAGAGACCGCGTCCGGTCTGCTCGCGCGCTGCCCGGGACACCGCGGTGTCGCGCTC
 GAATCCAGTTGGCGAAACGGATGATGGCGCTGCACGCGCTGATGGACGGCTATGGATCG
 GCGCGCTCGAGACGTCTCTCCAGGTACCCAGTGGGAACAGCTCTTCGACAACACC
 CAGCCACCGGGCTGCTGCACGTCCAGCGCTGAGCGCGCTCACCACGGGCGGCCATGAG
 30 CAGGCGGCGAGCTGGCGGGGCTGCTGCACGAGCTGGCCGACCACGGCGGCGGACCGGAG
 GCCCGTCCGGCGGCTGCTATGTGGACGGCGTCAACCCTGTATGTGGGCGGACGGGTGAC
 GAAGCCCTCGCCGCGCTCGCCAGGGCACCGAGATCACGGACGCCCTCCTGGCCGGACAC
 CGCAGGACCGCCGCCCGCACGGCGGCGGGCACCTCCAGGACCGGCGTATCGACTTCCGC
 GCTATCTGGCGCTCGGCCACTGTCTCAGCGCGACCGGATTAGACCCAGCGCTACCGG
 35 ACGGAATCCTCCACCTCACCCAGTCCGAACGGTACGACCGGCGGTGGGACCGGGCTTC
 GCGCGCTATGTGGACGCGTCACTCGCGTACCGAGTGCATGTCCAGGGGTGTGGCTG
 GCGCGCGGGCGGGGCTCGACCTCGCCGCCCGCTGCCAGTCCCGTCTTGGCAGCGGATG
 CTCGCGCTCCCCCTCGGTGGGCGGAGGTCCACCAGGGGGCGCACGACAAGGGGCTGGCC
 CGGATGCGGGAGGCGCTGCACGAGGCGGCCCGCACCGGACCCGTGCTGCGCCGTACGCTC
 40 CACCTCGGCCTGCTCGCCGACGCCCTCCAGTACAGGGGCGCCGGGAACAGGCCCGGCGC
 ACGATGTCTCCGCCGTACGGGAGATCGAGCGCGCGCGAGTACTTCTGTCTCCGGCCG
 CAGTGGCCCTGGGCGCGGCTCCTCCACAGCCACGGCACCTCCGCCGCGGCGGAGCACCGG
 GTCGTCCACGGCAGGCACTGACCCGGGGCGGCCGAGCCGGGCGGCGCCGACGGTTCGGGGC
 CGGCTCCGACCCGCGCGGCCGAGCCGGGCGGGCGGGCGGGCGGCCGACGGTTCGGGGC
 45 CGGCGGTGTGGGAGGGGGCGGCCCGGATCGCTCAGACCGGGCAGACGGCGGACCGCG
 CCGCGGCCGCGGAGCCGCGGCCCGCGGAGTCCCGTAGTCGCCCGCGAGGAAGAC
 CAGGGCGGAACCTCGCGGATCACCCGAGGTGCGGCACCGCCCGGTGACGAACAGTG
 GTCGCCCGCTCCGTCTCCCCGCCACCTCGCAGTCAACACGCGAGCGCTCGAGCAG
 GACGGGGGAGCCGGTGGCCGTGCTCCGTACGGCACCTCCAGCGCCCCGGATCGCCCCC
 50 GGCGAACTCCGGCAGACCGGGCCCTGATCCGCGCGGAGCACATTGACGCAGAAACGCC
 GGCGGCCGGAGCCGCGGCCAGGTCTGACGACCTGGCCGGGAGGAACCCACCAGCAC
 CGGATCGAGCGACACCGAGGTGAACGTCCCCACCACCATGGCGGGCGGCGGTGCCCGG
 AGCCTCGGCCGACCGGTGACCAGGACACCCCGGTGGGATAGTGGCCCGCCACCCGGCG
 55 CAGCAGACTCCCGGACACGGACCCGTGGGTGTGCGCGGAAGGCCCGGAGGCCGGGTAC
 AGCCACGGGTAACGCGCGGTGTCTTGGCCCGGTAATCGGGGTCCAGATAGACGAAGGCC
 CGGTGGACGAGGAAGTCCCGCACCTCGTAGACCGTGCACAGCGCCCGCGGCCACTCG
 GGGTACCCCGCCCGCACGGCCCGTCCCGGTGCTCACCGTGGGTGGTCCCTCCGCGCG
 AGGAGTTCCGTCCCGTCAGAATCCAGTTGACGGACACAGATGGTGGGTGATCGAGCGG
 60 ATGGTGGCCCCGAGGTCTGCAAGAGCCGGGCGATCTCGGACTGCCCCCGGGCCAGACCC
 CACTTGGGGAAGAAGAAGACCGCGTCTCGCGGAAGTAGTCGATCGCGGGGGTGGCGTGC
 CTGCCGACGCCCGGTTGTGCAACGCCTTGAAGTACCGGTGATGACCGCTTGGCTGC
 TCGTCCGTATACCGGCCGATGCCACGGACATGAAACGACCTCCAGAGATTCCGGGTGCG
 TGTGCTGGGGCTGCGGAAGGGGTGTCCCCCGGAAGGACGGCGGACGCCGCGGACGCCG
 65 GGCGTCTCCCCGGCGGACGGGTCCAGCGTCTTGAGAGGGGCTTGGCGGCGGCTTGACG
 CCGTGTGTCGCCGCGGCTTGCGGAACGCGAAGTACCGGCCAGCGTACGGGCGTTGACCG
 GACGTGTACGCCGGTCCGGACCCCTCGTACCCCGGAGCCGGCGGACCCCGGCGGCTCCG
 GGGGTACGGACGCGCGGACCGGCCGAGCGAGCGGTCGGACGGTGGCGGTGCGCGTGGT
 TCCGGTGTGTCGGACAGCTCGGACGGACCGGACGGTGCAGGTGGTTCCGGTGTGTCGGAC

AGCTCGGACGGGTTCGGACGGTGCGCGTGGTTCCGGCACGCCGGACGGGTGAGTTGCCGAT
 CATGGCGAGCAATGCCGGGGTGTACCGCTCCCGGACACCGGGTGGGAGATCGCGGCCGT
 CACCTCCGCGAGGACCGGTTCGTCCAGCCGGATCGAGGCGGCGGAGATTGTCCGCGAG
 ATGGGCCGGGTTCGCGGTGCCCGGGATCGGGACGAGTCTCGCCCCGGTGGTGACGCCA
 5 GGCAGCGCGAGCTGTGCCAGGGTCAGCCCCAGACCGTCCGCGACCGGGCGCAGCCGTG
 CAGCAACGAGCGGTTCGCGCGAGGGCCGGAGCGCTGAACCGGGGCTGGCCCCGGCGGAA
 GTCCTCGTCCCCCAGATCGTCGGTGGTGCGGATGGTGCCCGTGAGAAAACCCCGTCCCAG
 AGGGCGCTAAGCGACGATCCCGATCCCCAGCTCCCGGCAGACGGGCACCACCTCGTCTCTC
 GATCCCGCGCGACACAGGCTCCACTCGTCTGTGACCGCCGTACCGGGTGACCGCGTC
 10 CGCCCCGCGCAGCGTGCGCGCGAGGGCTCGGAGAGACCGAGCTGCGGACCTTGCCCTC
 GCGCACCAGCTCGGCCACCGCACCCACGGTCTCTCGATCGGCACCGCCGGTCCGTCCA
 GTGCTGGTAGTACAGGTTCGATGCGGTTCGTGCGGAGACGACGAGGGACCGTTCGAGGC
 CGCGCGGACGTAGGACGGCTCGCCGACAAAGCCCTGGGAGGCGCGTTCGGACGAGCGCAC
 CATGCGGAACCTTGGTGGCGATCAGCACCTCGTCCCGCGCGCCGCGACCGCCGTTCGAG
 15 CAGTCTCTACCGCGCGGAGCCCCGAGCGTTCGGCGGTGTCCAGCAGGGTGACCCCGGC
 GTCGACGGCGGCGCGGATGGTGGCCGTGCGCCGGGCGCGGTCCGGGCGTCCGTAGAAGTC
 GGTGGTTCGGCAGGACCGGAGCCCCGAGCGTGGGCACTGACCGGAAGGTCCCGCAGGGCGCGGAC
 CGGCGGACGCGGAACCGCGGCGGACACGGAACCGGCCGGGACTCGGGCGGAGAGCGGGA
 CATACGGAACCTCCACAGGCGGAGCCGGGAACGGGACGAGGCGAGGACGGGACGGAACG
 20 AAGGAGAGGACGGGACCGGACAGCAGGACGCGGACGAGACGGAACCGAGTTCGGGAACCGGG
 GGGGTGACCGGAACCGGGCGGTCTTGGCCCTCCCCCGTCTCCCCGCCATCCGCCGTTC
 TCCCCCGTTCCCTCTCCCGTCTCCAGCCAACACCGCCGCCCTTTCCAAGCGTTGACAC
 GGCACCGACAGCCGCGCCCGGGCGCCGATGGGGACCCGTGCCCGCCGGTGAGCGCGGT
 GAGCGCCGGTACGGGACCCACGCGCGCCCGCCGGGCGCCCGCAGGGCCCGCGCGGCC
 25 ACCCGGCGCCCGCCCGCGGAGCGCGGATCCGGGCGCTCGCTGCAAGAGGAACATCCA
 CAGCCGCAACAAGGAGCGCTCCGACAGTGGGCACACGTCCGCCCGTCCCCCACACCGT
 GGCCGTTCCCACCGGACAGCACAGCACCGCACAGCACCATCGCACGGCACAGCACAG
 CACCACCGGACAGGGAACCAAGGAAGGAACCAACACCATGACCTCAGTGGACTGCA
 CCCTGTACGCGCTTCTTGACGCGCGCGCACACAGCCCGCCCTCGCTCCCCGGCGCCCTCG
 30 CCACCGCGCTGGACACCTTCAACGCGGAGGGCAGCGAGGACGGCCATCTGCTGCTGCGCG
 GCCTCCCGGTGGAGGCCGACGCGACCTCCCCACACCCCGAGCAGCACCCCGCGCCCG
 AGGACCGTCCCTGCTGACCATGGAGGCCATGCTCGGACTGGTGGGCGCCCGGCTCGGTC
 TGCACACGGGGTACCGGAGCTGCGCTCGGGCACGGTCTACCACGACGTGTACCCGTGCG
 35 CGGCGCGCAACCTGTCTCGGAGACCTCCGAGACGCTGCTGGAGTTCCACACGGAGA
 TGGCTTACCACCGGCTCCAGCCGAACCTACGTATGCTGGCCTGCTCCCGGGCCGACCAG
 AGCGCACGGCGGCCACACTCGTTCGCTCGGTCCGCAAGGCGCTGCCCTGCTGGACGAGA
 GGACCCGGGCGCCGCTCTCGACCGGAGGATGCCCTGCTGCGTGGATGTGGCTTCCGCG
 40 GCGGGTGGACGACCCGGGCGCCATCGCCAGGTCAAACCGCTCTACGGGACGCGGACG
 ATCCCTTCCCTCGGTACGACCGCGAGCTGCTGGCGCCGAGGACCCCGCGGACAAGGAGG
 CCGTTCGCGCCCTGTCCAAGGCGCTCGACGAGGTACGAGGCGGTGTATCTGGAGCCCG
 GCGATCTGCTGATCGTCGACAACCTCCGCACCACGACGCGCGGACGCCGTTCGCGCCC
 45 GCTGGGACGGGAAGGACCGTGGCTGCACCGCTTACATCCGCACCGACCGCAATGGAC
 AGCTCTCCGCGGGGAGCGCGCGGGCGACGTGCTCGCCTTCAACCGCGCGGCTGAGCTC
 CCGGGTCCGACACCGCGCGGTGAACCCACGGTCCGGGGCCACGGTCCGGCACCGCGCG
 GCTGAGCCCCGGGTCCGGCAGCGGGCGGTGAACCCCGCCCGGGCCACCGCCGACC
 GCCCCCGCGACCGGACGCGCCCGCTGTACGGCGGTCCCGCCGGGCGCGTACACCTGA
 50 AGCGCCCGGCGGACCGCGCCCGCGCGGGGACGGAAGAGCCGGGTGCGGGAGGACGTC
 CTCCCGCACCCGGCTCCACCGTTCCGCACCGACCGCACCCGACCGTGCCGACGGCGCCA
 CGGCACCGCACCGCCCGCGCGGAGCCACACAGGCGCCACGCGCCCGCACGGTGGC
 CGCGTGTCTAGCCCCGTCCACCGGGTGTCCAGCAGCCGCGCGAGCGCGCCCGGATG
 AACTCCCGGTTCGGCGGCCGACCCCGGACCCCGGAGATGCCCCACACTCCCGGGATC
 55 ACCTCCAGCGAGGCATACGGCAGCAGATCGGCCACCCGCTTCTCGTCTCGACGGCGAAA
 CACACGTCCAGGGCGCCCGGCGAGCACCAGGCCCGCGCGGTGACGGAGGCCAGCGCGCC
 TCGACGCTCCCCCGGCCCCGGGTGTGCCCCACATCCGTGTTCTCCAGGTGCGCACC
 ATGGTGAGCAGATCCGCGCGCGCGGGCCCGGAGAGGAAGACCTGCTCCAGAAGCCGGTG
 AGGTACTCTCGCGGTGGCGAAACCCAGCTCCCGGTGGGCACGGCGGGCCAGAAAGGAA
 60 CGCGAGGTCCCCACCCGCGGAACACCCGCGCCCGCGCTTCCGCCCCGCTCCCCGGCG
 TCGGCGCTGAGCGCGCGGCCAGACCGGACAGCAGGACAGGCTGTGCGGGTGTCTACC
 GGCGCCCCGAGATCGGGGCGATCCGGCGCACCATCCCGGATGCGACACGGCCACTGG
 TAGGCTGGGCGCGGCCCATCGACACCCGTGACCGAGGCGAGTTCCCGTACCCCCAGC
 TCCTCGGTGAGCAGCCGGTGTGCGCCGCGACATTGTCTGCGGAGTGATCAGCGGAAAG
 65 CGGGACCCCGACGGGTGGTTGCCGGGCGAGCTGGAGACCCCGTTGCCGAAGAGTCCGGCG
 GTGACGACCGAGTACCGCCGGTGTCCAGCGGCAGCCCCGACCGATCAGCCAGTCTGTAC
 CCGGTGTGGTCCCGCCGAAGAAGCAGCAGGACAGAGCACCAGTTCTGTCGCTCGCGGTT
 GGCGTGCGGTACATCGGTAACCGATCCGGGCGTCCCGCAGGACCTCCCGTCCAGCAAC
 GGCAGTTCTGCTGATCTCGAATATGCGGCATTCCACCGCTGACCTCTTGTTCGATCCCC
 CGGACAACAGGTCCGTCTG GCCGAGACTCAGAGCCAGTTGGGGCGATCTCGGTGGCC

CACAGCTCCAGGCTGCGCAGCTGGACATCGTGCGGGATCAGCCCGAGTACTGGCACTGG
 AGCAGATACTCCGGATCGTGCCGCTCCACCAGCTTCTCGATCATGCGGTTGATGTCTGCC
 GGGGTGCCGACCCACTCCAGCCCCGGTTCGACAGGGTCTTGTAGTCCGAGCCGATCGGA
 5 CCCGTCTCGCCGGTGCAGCGCAGCGCCTCGGTGAAGCCCATGGGGCCGAACCACTTCTCG
 AAGATGAAGCCCGCCGCGCGGACGCCAGTGGTGGGCTCGCCGGAGTCCCGGGAGACC
 AGGACGTCTTTCATCACCCCGACCCGCTCGCCCCCGCGCAGGGTGGCCGCGCCGCC
 TCGGCCCTCTCCCGGTAGATGTCCATCAGCCGGGCGACGATCTGGTCTCGGTGTTTCATC
 AGGATCGGCACCCAGCCCTCCCGGGCACAGAACCAGGAACGTGTCTCACTGAAGCTGAAC
 10 GGCTGGAAGACGGGCGGGTGGGGCGCTGGTAGGGCTTGGGCGCGATGCCACCTCGCGG
 ATGACGCCGTTCTCGTCGAGGCCCGCGCCGTAGCGGCGCACCGCCTCGTAGGGGAACCTC
 AGGTCCGGCACCGGGATCGTCCACTGCTCCCCGGAGTGGGTGAACGTCTCGGTCTCCAC
 GCCTTCTTGATGATCTCCAGTGCTCTCGAAGAGGGCACGATTGCGCCGGTCCCGCTCC
 CCGGCGTCGGACAGGGTGGCCGCCACCCGTACACCTGCCCATGATGTGCGCCAGCGC
 15 TTCTGGAACCCGCGCGCATCCGACGAAGGCGCGGCCCGGGTATGTGGTTCGAGCATC
 GCCAGATCTCGGCCAGCCGAGCGGATTTGTGCAGCGGCAGGACGTTGGCCATCTGGCCG
 ACCCGGATGTGCCGGTCTGCATGCCGAGGTAGAGCCCCAGCATGATCGGGTTGTTGGAG
 ACCTCGAAACCTCGGTGTGGAAGTGGTGTCTGGTGAAGGACAGTCCCCAGTAGCCGAGT
 TCGTCCGCCGCTGCGCTGCCGGGTGAGTGCCTGGAGCATGTCTGGTAGTTCTGCGGA
 TTGACCCCCGCCATACCCCGCTGGACCTGCGCATGACTGCCGACCGTTGGCAGATAGAAG
 20 AGAATGGACTTCACCTGGCTCTCTCGGTTTCGCGGCGCCCTCCATTGACGTGCGCCGAAA
 GCGGCTCGACCCGTCCCACTCCGCCCTTGAGTTCCGTCTGACGCCGCGCCAGTGGCGGGC
 CGTCCGCCGGGGTGGCCGCCGGGGTCCGCACCCGCCGACGGCACGGCGCGCACCCGCGC
 CGCGGCGCTTCGGGGCACCGGCTCGACGGGGTCTCAGCGGGACGTCCAACGGAAGGCA
 AGCCCCGTACCCAGCCTGGTCAAGGCGCTCATCGCCATTCCCTGAGGAGGTCCCGCCTT
 25 GACCACAGCAATCTCCGCGCTCCCGACCGTCCCGGCTCCGGACTCGAAGCACTGGACCG
 TGCCACCTTCATCCACCCACCCCTCTCCGGAACACCGCGAAGGATCGTGCTGACCTC
 GGGGTCCGGCAGCCGGTCCGCGACACCGACGGCCGGGAGTACCTGGACCGGAGCGCCGT
 CCTCGGGTGACCCAGGTGGGCCACGGCCGGGCGAGCTGGCCCGGGTCCGCGCCGAGCA
 GATGGCCCGCTGGAGTACTTCCACACCTGGGGGACGATCAGCAACGACCGGGCGGTGGA
 30 GCTGGCGGCACGGCTGGTGGGGTGGAGCCCGAGCCGCTGACCCGCTTACTTTCACAG
 CGGCGGGCCGAGGGCAACGAGATCGCCCTGCGGATGGCCCGCTTACCACACCGGCG
 CGGGGAGTCCGCCCGTACCTGGATACTCTCCCGCGGTGCGCCTACCACGGCGTCCGATA
 CGGCAGCGCGCGCTCACCGGCTTCCCGCCTACCACAGGGCTTCGGCCCTCCCTCCC
 GGACGTGCACTTCTGACCCCGCCGACGCCCTACCGCCGGGAGCTGTTCCGCCGTTCCGA
 35 CGTCACCGACTTCTGCCTCGCCGAACCTGCGCGAGACCATCGACCGGATCGGCCCGGAGCG
 GATCGCGCGGATGATCGGCGAGCCGATCATGGGCGCGGTGCGCGCGCGGCCCGCCCGC
 CGACTACTGGCCCCGGGTGCGCGAGCTGCTGCACTCCTACGGCATCCTGCTGATCTCCGA
 CGAGGTGATCACGGGTACGGGCGCACCGGGCACTGGTTCGCCCGCGACCACTTCGCGGT
 40 GGTCCCGGACATCATGGTCAACGCCAAGGGCATCACTCGGGGTATGTGCCGCACGGCGC
 CGTCTGACCAACCGAGCCGTGCGCGACGAGGTGTCGGCGACCAAGGGCTTCCCGCGGG
 CTTTCACTTACAGCGGCCATGCCACGGCCCTGCGCGGTGGCCCTGGCCAACCTGGACATCAT
 CGAGCGCGAGAATCTGCTCGACAACGCCAGCACCGTTCGGCGCTTACCTGGGCAACGCCT
 GGCCGAGCTGAGCGATCTGCCGATCGTGGGGAGCTCCGGCAGACCGGTCTGATGCTCGG
 45 TGTCGAACCTGGTTCGCCGACCCCGGAACCCGGGAGCCGCTGCGGGCGCGCCCGTCCCGA
 GGCCCTGCGCGAGCGGGCGGGCATCTGTGTCGCGCCAACGGCAACGCCCTCATCGTCAA
 CCCCCGCTGATCTTACCCAGGAAGACGCCGACGAACTCGTGGCGGGCTGCGCTCCGT
 ACTCGCCCGCACCGAGCCGACGGCCGGGTGCTCTGACCCCTTTGGCCCTCCCGGCCCGC
 ACCGGGGCACCAACCCCGCCGACCCCGAGCGCAAAAGACCCCTCTGCCTGCGTTTCCGC
 50 AGGTGAGAGGGTCTGGTGCAGTGGAGCCTAGGGGAGTGAACCCCTGACATCTGCCATG
 CAAAGACAGCGCTTACCAACTGAGCTAAGGCCCCGAAGCGACAGAACGGCCCTGGACTG
 CTCCGTCCCGGCACTGCCGAGACAGAGTACCGGTGTTCCCGGTGATCTTCCAAAAC
 ATTGAGGTCTCCCGGTGGGCGACCACTCTCCGTAAGATGCTCGACGTGGTTCCGAGCAGC
 GAAGCCCGCTTGGGGAAGCGATGGGGAGACGCGCATGGACGCCGCTCAGCAGGAGACGAC
 55 CGAAGAGCCCGGAGCTACAGCGAAGCTGGTACGGGGAGCCCTGGGGGCCCTGTTCCG
 CAGGCTGATAGACGATCTGGGGCTGAACAGGCGCGTCTCGCGCGGTGCTGGGCCCTCTC
 CGCCCCATGCTCTCCAGCTCATGAGCGGCCAGCGGCCAAGATCGGCAACCCGCGCGT
 GGTCCAACGGGTCCAGCGCTCCAGGAGTTGGCCCGACAGGTGGCCGACGGCAGCGTCAG
 CGCGGTGGAGGCCACCGACCGCATGGAGGAGATCAAGAAGTCGAGGGAGGCTCCGTCTC
 60 GACCGCGAACAGCCAGACCAACAGCTCGGGGGCGCCGACCGTCCGCGGGGTCTCCG
 GGAGATCCAGTCTGCTGCTCGGGTCCGTGTCCGCCGCGGGGGACATCATCGACGCGCGAA
 CTCCCTCGCCCCGACCCATCCGGAGCTGGCAGAGTTCCTGCGGGTGTACGGGGCGGGCG
 CACCGCGGACGCGTGGCGCACTACGAGTCCCAACAGAGCTGACGACCGAGGCCGGCCCC
 65 GGAACGAGACAGAGCTCATGAGGGACGGGGAGCGGACGCGGCACCATGGGTGAGGTCTT
 CGCCGGCCCGTACGAGCTGGTGCACCCGATCGGACGCGGAGGGGTCCGCCCGGTCTGGCG
 CGCCTGGGACCAACGGCGCGCCGCTATGTGGCGGCAAGGTGCTCCAGCAGAGCGACGC
 GCACACCTGCTGCGCTTCGTCCGCGAGAGGCCCTGCGGATCGACCATCCCCATGTCCT
 GGCCCCGCGAGCTGGGCGCGGACGACAAAGTCTTCAACATGATCTCGTGGG
 CGGCGGATCACTCGCGCACGTGATCGGCGACTACGGCCGCTCCCGCCGCGCTATGTGTG

CGCCCTGCTGGACCAACTCTCTCCGGGCTCGCCGCGGTGCACGCCGAGGGCGTGGTGCA
 CCGCGACATCAAACCGCGGAACATCTTGATGAGGGCCACCGGGACGGGCGGCCCATCT
 GCGCCTGTCCGACTTCGGCATCTCCATGCGCAAGGGCGAGCCCCGGCTGACCGAGACCAA
 CTATGTCGTGGGTACGCCCGGTACTTTCGCCCCCGAGCAGGTGAGGGCGCGGAGCCGGA
 5 CTCCCCGCGGATCTCTTCGCGGTGCGCCTGGTCCGCTCTATCTGCTGGAGGGTCAGAA
 ACCCGACACCAAGGCCCTGGTGGACTTCTTCACCGCCCATGGCACCCCGGTGCTCCCCG
 GGGGATACCGGAGCCGCTGTGGCAGGTGCTCGCGGGGCTGATCCAGCCCGACCCCGCCGC
 CCGGTTCCGTACGGCGACGGGGGCCCGGAAGGCCCTCGCCGCGCGCGTGGAACTGCTTCC
 CGAGACGGCCCCGACGACGAACCGGTGGAGATATTCGACCAACTGGGCCCGCTGCCGCC
 10 GGGGTTCCGGCCCCGGCGGCCCGAGAACACGCCGCCCTCGGGTCTGCTGCGCTCGGCGG
 CTCCGGTACC

SEQ ID NO:18 orf2par reverse complement

ATGGCCACCACGACCGCGAAAGCCATGCTGGAACGCTTTCACAGTACGGTGTGACCATGTATTTCGGCGTCGTCGGCCGGGAGGC
 15 GTCGCCCATTTCTTTCGACGAGGTGGAAGACTCGACTTCGTCTGACCCGGCACGAGTTCACCGCCGGGGTGATGGCGGACGTCC
 TCGCCCGGATACCAACCGCCCCAGGCGTGCTTCGCGACCCCTGGGCCCGGCATGACCAACCTGGCCACCGGCGTCGCCACCTCC
 GCCCTGGACCGCAGCTCGGTTCATCGCGCTGGCCGCGCAGTCCGAGTCTGACGACTGCTACCCCAACGTACCCACCAAGTGCCTGGA
 CAGCACCGCGGTGATGGGCCGCTGACCAAGTTCAGCGTCCAGCTCGAACCGCGCGAGGACATCGTCAACCTCGTCGACAGCGCCG
 TCCTCAACAGCCGGATCGAGCCGTGGGTCCAGCTTCATCAGCCTGCCGTCGACCTCCTCGGCGCGGAGCTGAACGGCACCCCC
 20 ACCGACGCCCCCTGGTCCGGGCCACCGCCACCCACGCCCTGGACGCCGACTGGCGCGCCCGCCTCGACGAGGCGCTGAGCTGGT
 GCGCGAGGCCGAGAACCCCTCCTCGTCTCGGTAGCGCGTTCATCCGCGCGGGGCGGTGACGCCCTGCGCGCCCTCGCCGAGC
 GGCTGAACATCCCCGCTGTCACCACTACACCGCCGAAGGGCGTCTGCCGACGACCAACCCGCTCAACTACGGCGCCATCAGCGGC
 TACATGGACGGCATTCTCGGCCACCCGGCCCTCGACGAGATCTTCGGCCCCCGCACCTCCTCCTGGCGATCGGCTACGACTACGC
 CGAGGACCTGCGCCCTTCATGTGGACGCGGGGCCGGGCCAAGACCACGGTCCGGGTGCGCCCGGAGGTCAACCCGATCCCGGAGC
 25 TGTTCCGCGCCGACATCGACATCGTCACCAACGTGCGCGAATTCTGTACCGCGCTCGACGACGCGACCTCGGGCCTCGCCCCAAG
 ACCCGGCACGACCTCAGCGCCCTGCGCGCCCGCTGCGCGAATTCTCGCCGACCCACCGAGTACGAGGACGGCATGCGGGTCCA
 CCAGGTGATCGACTGCATGAACCTCCGTCTCGACAAACGGCACCTTCGTGACGACATCGGCTTCTTCGCCCACTACGGCGTCTCT
 TCGCCAAGTCCGACGAGCGGTACGGATTCTCACCTCGCGGGCTGCTCCAGCTTCGGCTACGAGTGCCTCGCGCCGATGGCCGCC
 30 CAGATCGCCCGCGCCGGCGAGCCGCTTCTCTCATCGCGGGCGACGGCGGCTTCCACTCCAACAGCGCCGACATCGAGACGGCCGT
 GCGCCTGGGCTGCGCATCGTCATGGTCTGTCGTAACAACGACCGCAACGGCCTGATCGAGCTGTACAGAACCTCGGACACCAGC
 GCTCCCACGCCCCCGCGTCCGGTTCGGAAGCGTCGACTTCGTCCAGCTCGCCGAGGCCAACGGCTGCGAGGCCGTCCGCGCCACC
 GACCGCAGCTCGTCTGTCGCCGCCCTCACCAAGGCGCGGACTCGGCCGCCGTTCTGATCGAGGTACCGGTGGCCTACGACTT
 CCAGTCCGGCGGTTTCGCCGCCCTGGCCATCTGA

SEQ ID NO:19 orf3par reverse complement

ATGCCCGGCCCCGACCTCGTGTACGGATTCCGGGTGCGCATCGGCACCGAGGGCCGCCCGCGCGGCCCGCGGTCCTCCGA
 ACCCGGCAGCGCACCCCGCTTCGCGCTCCGCGGACCCATGTCCCCGTGACAGCGGCACCGCGTACCCGCTCTGGAGCGGAACGG
 35 CCGTGACCTGGGCGCTCGCCCGTCTGGTTCGCCGACCGGCCAGTTCGGGTGCTCTCGCGGGCGAGCTGTACAAACCGCGCCGAG
 CTGACCGGAGCGCTCGCGGCTCCTCTGCCGCCCTCGGCGACGCGGAAGTGTGCTGGCCGCTGGCGGCGTGGGGCCCCGGGGC
 40 CTTCGGCTCCTGAACGGACGGTTCGCCGCACTGCTCACCGACGCTCCACCGGCGCGACCGTTCGCGGCCACCGACACGCCGGTT
 CGGTACCGCTGTGGCTGCGCGCCGACGTGACGGGCTGAGCGCCGCCACCGAGGCGAAGACCTGGCGCACGAGCCGGGCCGGCCG
 CTGGGCTGTCCGGCACCCACACCCGCGGGGGCGGGGCGTTCGCCGGTCCCGCGGGGACCGCCCTCCTGCTGCACGGAGTC
 GCGGCTCCGACATCACCGCAGGGCGGTCCGACCTGGACACCCCGCTCTCCGGGCGTGCCTCGGCGAACCGGAGGCGGTGGA
 45 CCTGTCGGCGGAACCGTCCGACGCGGCTCCGACCGGCTGCGCGCGGGGAGGCGGCCCCACCGTCTGCTCTGTCGGCGGCA
 TCGACTCCGGGGAGTTCGCCGCCACACGGCGGCCCTGGCACCCGGGACACGGTCCGTGTCGATGGGCACCGAGGTGTCCGACGAG
 TTCGACGCGGCCCGCTCGGTTCGCGTCCACCTGGGCACCGCGCACAGCGAGATCCGGTCCACTCGGCCGAACCTCGTCAGGGAAC
 GCCCTGGGCGGTTCGCGCGCGGAGATCACCGACCCACGGTCTGGAGTACCTGCTGCCGCTCGTCGCCCTTACCGGCGGCTCG
 50 ACACGGGCGGCTCCGATCTCACCGGTACGGCGCGGACATCGGCTCGGCGGTATGACACCGCGCACGCGCTCGTCTGGTCC
 CTCGACGACGAGATCGCGGGCGACATGGCGGGCTTCGACGCGCTCAACGAGATGTCCCCGTCTTCGCGGGCATCGCCGGGAAGTG
 GACCACCCACCCGTAAGTGGACCGCGCGGTCTGGACGCGTGGTCTCCCTCGAACCCGGGCTCAAACGCGCGGGGCGGCCGACA
 AGTGGGTGTTCGGCGAGGCCCTCTCCGGCTGTGCGCGCGGAGACCGTGGCCCGCCCAAGCTGGGCATCCACGAGGGGTCCGGC
 ACCACGACCGCTGGACCGGACTGCTCCTCGCCGAAGGATCCGGCGCGACGAGGTGACGGCGCTCAAGGGCGCCATGGCACGGCG
 CCTGTACGACGCGGTGGTTCATCGACACGGTGCGCCGCGGAGGACGTGGACTTCGGCGAGACGGTGCGGCGCTCCGTGACGCGGTG
 55 GCAGGCTCAGGTCCAGGGCCGGTGGTCTATGA

SEQ ID NO:20 orf4par reverse complement

GTGTCCACCGCCGTCTCCCCGCGCTACGCCCAACCGGCGACCTTCATGCGGCTGCGCCACCGGCCCGACCCGATCGGCCATGACGT
 GGTGGTCTGTCGGCGCCCCGTACGACGAGGACACAGCTACCGGCCCGCGCGCGTTCGCGCGCGCGCCATCCGGCAGAGTCCA
 60 GCCTGATCCACGGCGTCGGCATCGACCGGGGCCAGGGTCTTCGACCGGATCGAGTGGTTCGACGGGGCGACATCGACTCAGC
 CCGTCTCGATGGATCGCGGATGGACACCGCGAGTGGCTGACCCGGCTCTGGAACGCAACGACGCTTCTGATGCTGGG
 CGGGGACCACTCGCTCTCCCTGGCCGCCCTGCGCGCGGTGACGCGCCGACGGCCGGGTGCGCGTCTGCACTGGACGCGCACA
 GCGACACCAACCCACCGTCTACGGCGGCACCTACCACACGGCACCCCTTCCGCTGGGCCATCGAAGAGGGCCTGGTGGACCCG
 GAGCGCTGGTCCAGTTCGGCATCCGCGGCCACAATCCGCGGCCGACTCCCTGGACTACGCGCGCGGGCACGGCGTCAGCATCGT
 65 CACCGCCGCGACTTCACCCGGCGCTCACCGCGCGGATCGCCGAGCAGATCCGGCGCACCGTGGCGGCGCTGCGGCTGTACGCT
 CCGTCGACATCGAGTCTGACCCGGCGTAGCCCGGGGACCGGCAACCGGCCCGGGCTGTCTTCGCGGAGGTGCTG
 ACCCTGCTCGAGTGGTGGGCGAGCTCAGGCCGTGGCTTCGACGTGGTTCGAGGTGTCGCCCGCGTACGACCCGTGGGGATCAC
 CTCCTGCTGGCGCGGAGATCGGGGCCGAAGTCTTACAGTACGCCCGCGCCACCGTGCCTCGCGTGGCACCGGTGGACT

CTCCCCCTGCCACCGGGGGCGGCGGCGGACGACGCCGAGAACGCCGAGAACCGCGGTGGACGCGGTGGACGCCGAGAGCGCCGTGGAC
TTCGCCGGGCAGCGGTGGGGGTAG

SEQ ID NO:21 cvm6 Polypeptide

5 VPGSGLEALDRATLIHPTLSGNTAERIVLTSGSGSRVRDITDGREYLDASAVLGVTQVGHGRAELARVAEQMARLEYFHTWGTISND
RAVELAARLVGLSPEPLTRVYFTSGGAEAGNEIALRMARLYHRRGESARTWILSRRSAYHGVGYGSGGVTGFPAYHQGFPSLPD
FLTPPQPYRRELFAGSDVTDCLAEELRETIDRIGPERIAAMIGEPIMGAVGAAAPPADYWPRAELLHSYGILLISDEVITGYGRTG
10 HWFAADHFGVVPDIMVTAKGITSYVPHGAVLTTEAVADEVVGDQGFAGFTYSGHATACAVANLNDI IERENLLDNASTVGAYLG
KRLAELSDLPIVGDVRQTGLMLGVLELVARGTREPLPGA AVAEALRERAGILLRANGNALIVNPPLIFTQEDADELVAGLRSLARTR
PDGRVL

SEQ ID NO:22 cvm3 Polypeptide

15 VTRPPGLSAHNGSVSGSLRRRVAGHYPTGVVLVTGPAAEPGQPPAMVVGTFSTVSLDFVLVGFPLPARSSTTWPR
LRAAGRFCVNVLGADQGPVCRSFAGGDPGRWEVPTTATGSPVLLDALAWFDCEVAGETEAGDHWFTVGAVRDLG
VIREGSPLVFLRGDYGHWAGGGGSGRAGRRAVCPV

SEQ ID NO:23 orf6par Polypeptide

20 MRASSPRGFRVHHGAGIRGSHADLAVIASDVPAAVGAVFTRSRFAAPSVLLSRDAVADGIARGVVVLSGNANAGT
GPRGYEDAAEVRHLVAGI VDCDERDVLIASTGPVGERYPMSRVRAHLRAVRGPLPGADFDGAAAVALGTAGARPTI
RRARCGDATLIGVAKPGTGPAEQDDRSTLAFFCTDAQVSPVVLDDI FRRVADRAFHGLGFGADASTGDTAAVLAN
GLAGRVDLVAFEQVLGALALDLVRDVRDSCGGALVTVRVGTGAHDTAQAGRVGRAVVDAPSLRAAVHGPAPDWAP
VAAVAGGHGDEGPGRSPGRITIRVGGREVFAPPRDRARPDVA TAYPHGGEVTVHIDLGVPGRAPGAFTVHGCDDL
GYPRLGAGRAV .

SEQ ID NO:24 orf4par Polypeptide

25 VSTAVSPRYAQPATFMRLRHRPDPIGHDVVVVGAPYDGGTSYRPGARFAPRAIRHESSLIHGVGIDRGPVGFDRID
VVDGGDIDLSPFSMDLAMDTATVALTRLERNDAFLMLGGDHSLSLAALRAVHARHGRVAVLHLDHSDTNPPVYG
30 GTYHHGTTPFRWAIIEGLVDPERLVQVGIRGHNPRPDSLDYARGHGVSI VTAADFTRRSRPGIAEQIRRTVGGLPLY
VSVDIDVVDPAYAPGTGTAPAGGLSSREVLTLDDVVGQLRPVGFVVEVSPAYDPSGITSLLAAEIGAELLYQYAR
ATTSPASAPVDSPLPPGAAADDAENAENAVDAVDAESAVDFAGQRWG .

SEQ ID NO:25 orf3par Polypeptide

35 MPGPDLVYGFVRIGTEGRPGGGPGGHSEPGSAPRFVARGTHVPVHDGTAYPLWSGTAVTLGRPPVLVADGQVRLL
LAGELYNRAELTGALGGSSAALGDAELLLAAWRRWGPGAFRLNNGRFAALLTDASTGATVAATDHAGSVPLWLRAD
VTGLSAATEAKTLAHEPGRPLGLSGTHTAPGAAGVCRVPAGTALLLHGVGGS DITARAVRTWTPPLSRALPGEREA
VDLVGERLATAVRTRLRGGEAAPT VVLSGGIDSGGVA AHTAALAPGTRSVSMGTEVSDEFDAARS VAVHLGTAHSE
IRLHSAELVRELPAWAAAAEITDPTVLEYLLPLVALYRRLDTGPLRI LTGYGADI PLGGMHRRRTASLWSLDDEIAG
DMAGFDGLNEMSPVLAGIAGKWTTHPYWDRAVLDAVLSLEPGLKRRRGTDKWLRLQALSGLLPAETVARPKLGIHE
40 GSGTTSANTGLLLAEGIRRDEVTA VKGAMARRLYDAVVITVPPEDVDFGETVRRSVDAVRRLRLQGRVVV .

SEQ ID NO:26 orf2par Polypeptide

45 MATTTAKAMLERLHQYGVVDHVFVVGREASAILFDEVEGLDFVLTRHEFTAGVMADVLARI TNRPQACFATLGPGM
TNLATGVATSALDRSSVIALAAQSESYDCYPNVTHQCLDSTAVMGPLTKFSVQLERGEDIVNLVDSAVLNSRIEPV
GPSFISLPVDLLGAELNGTPTDAPLVRATATHALDADWRARLDEAAELVREAENPLL VVGSAVIRAGAVDALRALA
ERLNI PVVTTYTAKGVLPHDHPLNYGAISGYMDGILGHPALDEIFGPADLLLAIGDYAEDLRPSMWTRGRAKTTV
RVAPEVNPIPELFRADIDI VTNVAEFVTALDDATSGLAPKTRHDL SALRARVAEFLADPTEYEDGMRVHQVIDCMN
SVLDNGTFVSDIGFFRHYGVLFKASDQPYGFLTSGAGCSFGYGLPAAMAAQIARPGEPVFLIAGDGGFHSNSADI E
TAVRLGLPIVMVVVNNDRNGLIELYQNLGHQRSHAPAVGFGSVDFVQLAEANGCEAVRATDRTSLLAALTKGAGLG
50 RPFLIEVPVAYDFQSGGFAALAI

Claims

1. A *S. clavuligerus* microorganism comprising DNA corresponding to one or more open reading frames essential for 5S clavam biosynthesis, wherein said open reading frames are disrupted or deleted such that the production of 5S clavams by said *S. clavuligerus* is reduced and clavulanic acid production is at least maintained, wherein the open reading frames are selected from:
 - a) *cvm6para* (SEQ ID NO:1);
 - b) *cvm7para* (SEQ ID NO:2);
 - c) *cvm6para* and *cvm6* (SEQ ID NO:5); or
 - 10 d) *cvm7para* and *cvm7* (SEQ ID NO:6).

2. A *S. clavuligerus* microorganism comprising DNA corresponding to one or more open reading frames essential for 5S clavam biosynthesis, wherein said open reading frames are disrupted or deleted such that the production of 5S clavams by said *S. clavuligerus* is reduced and clavulanic acid production is at least maintained, wherein the open reading frames are selected from:
 - a) *cvm6para* and one or more of *cvm1* (SEQ ID NO:7), *cvm2* (SEQ ID NO:8), *cvm3* (SEQ ID NO:9), *cvm4* (SEQ ID NO:10), *cvm5* (SEQ ID NO:11), *cvm6*, *cvm7* or *cvm7para*; or
 - b) *cvm7para* and one or more of *cvm1*, *cvm2*, *cvm3*, *cvm4*, *cvm5*, *cvm6*, *cvm7* or *cvm6para*.

3. An isolated polynucleotide comprising open reading frames selected from the group consisting of:
 - a) *cvm6para*;
 - b) *cvm7para*;
 - 25 c) *cvm6para* and *cvm6*;
 - d) *cvm7para* and *cvm7*;
 - e) *cvm6para* and one or more of *cvm1*, *cvm2*, *cvm3*, *cvm4*, *cvm5*, *cvm6*, *cvm7* or *cvm7para*; or
 - f) *cvm7para* and one or more of *cvm1*, *cvm2*, *cvm3*, *cvm4*, *cvm5*, *cvm6*, *cvm7* or *cvm6para*.

4. An isolated polynucleotide comprising one or more open reading frames encoding one or more enzymes involved in clavulanic acid biosynthesis wherein said open reading frames are selected from the group consisting of:
 - a) *orf2para* (SEQ ID NO:12),
 - b) *orf3para* (SEQ ID NO:13),
 - 35 c) *orf4para* (SEQ ID NO:14), and
 - d) *orf6para* (SEQ ID NO:15).

5. An isolated polynucleotide comprising one or more open reading frames encoding one or more enzymes involved in clavulanic acid biosynthesis wherein said open reading frames comprise one or more of:

a) *orf2para*,

5 b) *orf3para*,

c) *orf4para*,

d) *orf6para*

in combination with one or more genes involved in clavulanic acid biosynthesis selected from *orf2*, *orf3*, *orf4*, *orf5*, *orf6*, *orf7*, *orf8*, *orf9*, *orf10*, *orf11*, *orf12*, *orf13*, *orf14*, *orf15*, *orf16*, *orf17*, or
10 *orf18*.

6. An isolated polynucleotide selected from the group consisting of

a) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:16;

b) a polynucleotide having the nucleotide sequence of SEQ ID NO:16;

15 c) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:17; and

d) a polynucleotide having the nucleotide sequence of SEQ ID NO:17.

7. A vector comprising the polynucleotide of any one of claims 3 to 6.

20 8. A *S. clavuligerus* microorganism comprising the vector of claim 7.

9. A process for improving clavulanic acid production in a suitable microorganism comprising isolating the polynucleotide of any one of claims 3 to 6, manipulating said polynucleotide, introducing the manipulated polynucleotide into a said suitable microorganism
25 and fermenting said suitable microorganism under conditions whereby clavulanic acid is produced.

10. A process according to claim 9 wherein the polynucleotide is a *cvm* or *cvmpara* polynucleotide and the manipulation comprises disrupting or deleting *cvm* or *cvmpara* gene
30 sequences.

11. A process according to claim 9 wherein the polynucleotide is an *orf* or *orfpara* polynucleotides and manipulation thereof comprises insertion of the polynucleotide into vectors suitable for expression.

35

12. A process according to any one of claims 9 to 11 wherein the suitable microorganism is *S. clavuligerus*

Abstract

New processes for improving the manufacture of clavams e.g. clavulanic acid. Novel DNA sequences and new microorganisms capable of producing increased amounts of clavulanic acid are also disclosed.

Figure 1. Diagram of the paralogue cluster

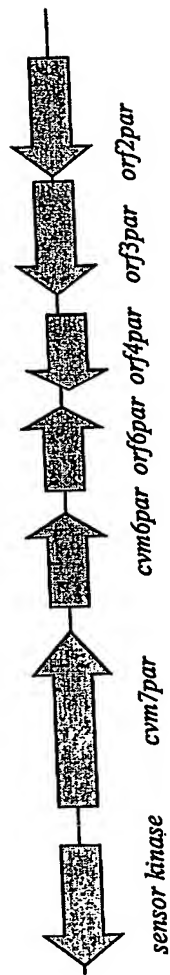


Fig.2 Orientation of *cvm7* to published *cvm* cluster

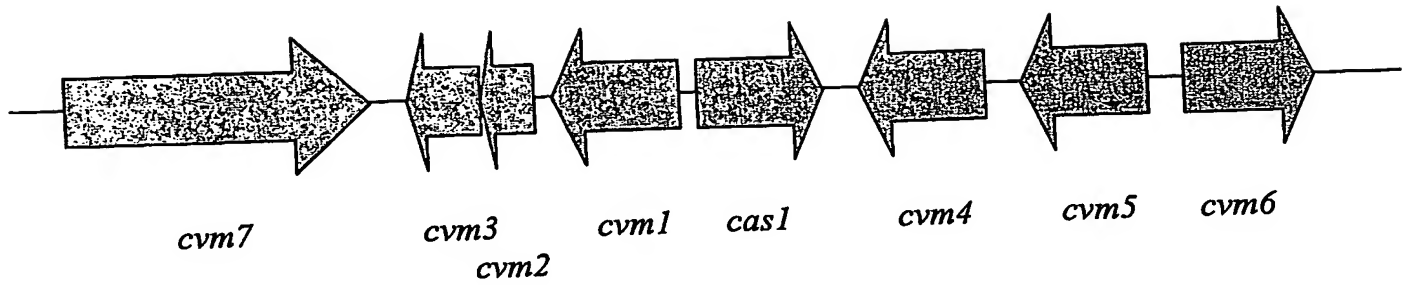


Fig 3. Annotated sequence of the paralogue cluster

```

      | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
1  ccatcgggagc agcatgcag tgcgcctccc cggcgcccat gccgtagct ggtagtcccc ctgcccgggtg ccgaccgcgcg 80
      | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
81  gggcggtccc ggggtcgccg gccggatcta gtcggtgtgc tccgacggtg cctgtcgggt gaggggcagt gtcaggcgga 160
      | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
161 tgggtggttcc cgcgcgcggc gggctgtgca gccgcagttg gccgcgagt gctccacccc ggtcgggtgag gccgacgagg 240
      | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
241 cccgagcccc ggcagggggc ggcgccaccg cggccgtcgt cgcggatgcc gacgtggagc cgtccgtccc gggtagccac 320
      | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
321 atggacgtcg acgacggtgg caccggagtg cttggcggcg ttggtcaggg cctcggagac ggcgtagtac gggcggtct 400
      | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
401 cgaccggttc ggggtggcgt tcccgggtct ggaatgtcag ccggaccggg atggcgagc gccggggcag ggccttgagc 480
      | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
481 gccggggcga gtccgccttc ggcgagtacc gccgggtgga tgccccgggc gacctccgg agttcgtcga cggcggcggc 560
      | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
561 cagcccgctg gtacactcgt cgagctgccg gatcagctcg tcggcgctca gcggcacca cagttgcacg gtgcgcaccc 640
      | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
641 gcagcgccag ggagaccagg cgctgttggg ggcgcgtcgtg caggttcgct tcgatacggc ggcggggcgt gtccggggcg 720
      | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
721 ggcagatcc gggcccgtag cgcggtgagg gccgcctgcg tctccgcgtt ggcgatggcg gtggccacca gttcgtgtaa 800
      | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
801 gccgggccag cggtcctcgg tgtccgacgg catcgggttg tegttcacg acgcacgct gagcgcgcc cacagttgtc 880
      | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
881 cgtcgaagtt gatcggcatg cacaccgtgg cgcgggaatcc ccactccttg ccgacgaagg aggcggggcc cgaaggacacg 960
      | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
961 gccgcgtagt cgtcgatccg cgcggggcag ccgactcga acaccagggt gtgcacatc cggccgcggc gcggtacctg 1040
      | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
1041 gataccggcg ggaataatcac ggcggtcct ggtccaggcg gcgacataca gggcggttcc gttgggtcg taacgggcga 1120
      | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
1121 ggaaccggaa gtcggccgag aggagtgtc cggcctcggc ggcgaccggc gcgaacacct ccttcggcg tgcggccgcg 1200
      | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
1201 gcgaccaggg tcgcacgcg ccgcagcgcc gcctgctcct cggcgggccc ccgcagctcc acacgtgcct ggggtgttcg 1280
      | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
1281 gatggcggtg gccacgaggt cggtgaaacc ggcacgcggg tcctcggtgt cggggcgag cgttcccgcg gtcagcgaga 1360
      | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
1361 tcgcacatcat cagccccac agcgttcctt cgaagttgat cggcacgcgg acgacggaac cgaagccgcg cgccctggcg 1440
      | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
1441 AAGTCGGCGG GTGCCCCGGA CGACTCGGCG GCGTCGTCCA TCCGGGCCGG CCGCCCCGTC TCGGACACCA CGGTACCCAC 1520
      | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
1521 GTTCCGGCCG TCGGGGTCCA CCCGGGTGCC GATGGGAAG AGCGGCCCGT GCAGACTTCT GGACACAGCG CCGACGGCCG 1600

```

1601 TCGCCATGCC GTCCGGATCG AGCCTGATGA TTCCGGTTCAC ATCTTTGCGG AGCAGTTCTC CGACTTCGGC GGCACCCGTC 1680
 1681 GCGACATCT GTTCCGGTGG GGTGGCCCTG GCGACACAGG TCGCCACCGG TCGGAGTGCC GCGCGTCTCT CGACGATCTG 1760
 sensor kinase ←
 1761 TTGACACGAC ACGACCGCTG CCAGGCCCCC CTTACCCGCC GATGACGCC GCATACCGGG TATCACGGCA CATCAGCATG 1840
 CGGGACACAG CTTCCGGGAT CAGCCCCCGG AACGGCGGGA 1920
 1841 ACGTCCGCGG TGAACCCCG TCAACGTGGC CCGCCGGAGT GAATCCGCC TGAACCTCGG AGTTTGACAG TAGCTGGAAT 2000
 1921 CCGTCTTCTT CCGTCCGGCG CGGGGCACAG CGCCCGCGCG GGCGATGCGG AAGCCGATCG TTCCCAGTAC TTCTGGGAAG 2080
 2001 CAGCGGTTCT GGTGTGGTGG AAGGATGTT GGCCGCTGGC GACGACGCTG CCGGTCTTCC ACGGGGGAGA GATCCGCGAA 2160
 2081 TGCGTCGCGG AGAGTCGGTC CGCTTCCCCG AGTGGGCCCG GAAGAGTTTC CCGGTGTTC GACCCCGCGC GGCAACGTCC 2240
 2161 CCGGCGGAGG AGTGCCTG TGCGACGTCT TCGCATCCGA ACGCAGAGAA GATCCGATAC GCAGTTGACG AGTGCAGCGA 2320
 2241 CCACCGCGCT CTGTCAATCA GCGCGTCGCG GCCGTGAGCC CCGTCCATCA CCGTCTTGGG CTGTCTGGGC GTACGCGCGG 2400
 2321 TGAGTTCTGT CACGACGTCC CCGGCTTGGC GGTCTCGTCA CGTTTTCGCC CTGTGTCTCA TCAACGCGGG CAGTGTGGTG 2480
 2401 ACGGCGGAA ACTGGAGCTG GGCCCTCCGC GTACAGCGGG CACCGGCGC ggteacgag acgtccagt cctatgtgtc 2560
 2481 CCGGTCGACT cgategttt ccgtatctgg ggaactcac ggttcgacac ccgaactgct gacacagcgg cggggtctaca 2640
 2561 cgggtgctgg aaactcctgg ccgagtgtgt gctccgggac tgagcaggcc atcaggacag ggcgcgggct ctgcgcgag 2720
 2641 cctcgcgct cggcaccgag cacatogaag cgaacogttt tgtgagctg gggcgggaca ccgtacgag agctgagcgc 2800
 2721 gagcagcacc agggggcgg ggcgtgtctc tgccaggccc ctccggctgg gcgcgtgga gacatgggag cactgtgttc 2880
 2801 gtaagacttc gccgtccagg aggcgaatcg gctggagcag GCGGAGGTG CAGCCCAATC CGCTGCGGGA GCGCTGATC 2960
 2881 tggggtggg ggcggacgag gagggtgatgg ACCAGTCAA CGGACGCGCT CAGGACGTAC GAGGCGACGC GCGGGGCCCT 3040
 2961 GGGCAGCTCA TGACGGCGCA GTACCGGCTG GGTGTCCAGG GCGCTGACAG CCGCGATCCT GCGTCAGGAC AACGCTCTGG 3120
 3041 GGCAGAGAG CTGGGAGACG ATCCGGGCAA GGAGTGGCG CGGGCGGGGG GCGGTGACGG TGTCTGGTCCC GGCACAGCGG 3200
 3121 ACCGCGTCTT CCGGCGTCC GCGCCGCGCT CCGCGGGGT TCCCGGGGGC GATGACGTG GCGGCGGGCG CGGGGGCGGC 3280
 3201 TCGAGGCCGT TGACGCGGCC GGTGGCGGG CGGGCGCGGG GGCTCCGGCT CCGGCTCCGG CTCCGCTCCT GCGTCGGTTC 3360
 3281 CCCCGCTCC GCTCCGGCT CCGTTTCGC GTCCGTTTC TCCCGGCTCC GTAGCGCGCG CCGTTTCCGG CCATGCTCTC 3440
 3361 CCACCTTCTT TCCCGGCTCC GTTCTTGGCT CCGGCTCCGT

3441 GGGCCCCGGT CCGCTTTCCG GTCCGTGGCG CTCCACCGGC CGCAGACCTT CCGGGCGGAG CCGGTCCACG GGGGCGCGCA 3520
 3521 GGGGATGCGC ACCGGCAGG TGTTCCTCAC GCTGCCGCCG TTGCTCGGCG GCGCGACGA GCTGCGGGT ctgctggagt 3600
 3601 ccgagacgtc cgggttccac acctcggggc gggtagcggt cgtcgtcggc gaggcgggca gggcaagac cgggtccctc 3680
 3681 tccagagttgg aggcctcggg tccggacagt gtgcacacg tctggggctc ctgttcggag agtgaggacc gggccgacta 3760
 3761 ctggccgtgg acgacgtgc tgcggcatct gtacgcatg tggccggaac gtatgcacgg attccccgtt tggctgcggc 3840
 3841 gcgcaactgc ggaactgctt cccgaggtgg gcccgagac acagggggcg cactcccccg acggggggga ggaacaacgc 3920
 3921 ggcacacggg acggtgcggg cgaaggggac agcacccccg cgcacacct cagctcgcg cccgtctcgc cgcccccgcg 4000
 4001 ctccagagag gctcgtttca cctgcacga cgcgtgtgc caggcgcttc tggcacagg cgcgaaccc gtggtgatca 4080
 4081 tgcctggagg catggagcgg gccgacgcc cctcgtcgc cctgctgcgc ctctgtgtgg agcaactgcg caccgtcccc 4160
 4161 ctgctgctcg tggtaaccac gcgaccttc cggctcgcg acgacgcga gctgcgacgg gccgcgcgcg tgatcctcca 4240
 4241 gtgcacggcg gcggcgccgg tctgtgtgaa cgccttgac gcacggggca ccggggaact cccggagggg atgctgggca 4320
 4321 agggccccga caccctctc gtacggggcc tgcacgagcg ctccgcgggg aaccgtact tctctgtcca gctcctcgcg 4400
 4401 tgcctcggcg aggggctcgc cgcgcgtgc gccgggtgct cccggacga gctggccggg gtctgtctgc aacggctgtc 4480
 4481 gagcgtgcg cccgcgtgc gccgggtgc cgaacatctgc gcggtcgtgg agcgagttg cgaacggcgt gtgctcgaga 4560
 4561 ccgtgctcgc ccatgaggga atcccgctgg agaactccg tacggcggtc cgcggcggtc tgctggaggga agaccccgac 4640
 4641 gacccccggc ggctgaggtt cgtgcatccg ctggtccggg aggcgtctc ggacgacctg gagaacaccc gtccggccgt 4720
 4721 gtcccgttcc tccggtctcg gggcgctggc caccgtctga gtcccgggcc cccgggtcct cgcgcgcggg CCGCGCTTGC 4800
 4801 GCGCTCCCGC ACGCCGGGCT TGATCCCCCG GGGCAGCCCG ACGCGAGCC GGGTGCAGG GCGGTGCGG ACACGTGGCG 4880
 4881 GCGCGCGGCG GTGGCCGGTC GCGCCCCCCC ACGGCCACC GAGGAGCCCC CATTTGACAC GTACGCAGCG GATACGTACC 4960
 4961 CCGGTTCCCG CACCACACCC GAGCGCGCTC CCGACGCACC TCCCCACGCG CGTCCGGGA CCGTCCCGG CCGTCCCTGC 5040
 5041 GAGCGCGGCC CCGACCCGGG CCGCAGGCC GCGTGGCTGC TCCTGCGGGA CCGCGCCCAT ATGTTCCACC CCGTCCCTGC 5120
 5121 CCGGGGCGCG GAGGACGCA CCGTTCTGGT CTCGGCCCGC GGCTGCACCG TACGGGACAC CGAAGGGCGC ACCTATCTCG 5200
 5201 ACGCCTCTGC GTGTCTCGA CTGACCCAGA TCGGCCATGG ACGTGAGGAG ATCGCGCAGG CCGCGCCCGA GCAGATGCGG 5280

5281 ACATCGGTC ACTTCCACAC CTGGGGCACC ATCAGCAAGC ACAAGGCCAT CGACTGGCC GCGCGCCTCA CGACTGGC 5360
5361 GCCCCAGGTT CTCCAGGCG TCTACTTCAC CAGCGCGGC GCGAGGGCG TCGAGATCG CCTGGGCATG GCCGTTACT 5440
5441 TCCACACCG CACCGGCAGC CCGAGCGCA CTGTATCTT GTCGCGCGC ACGGCTTACC ACGCATCGG CTACGCGAGC 5520
5521 GTTACGGTGT CCGGCTCGCC CGCTTACCAG GACGGTTGG GCCCGTGTCT GCCCATGTG CACCACCTCA CGCGGCCCGA 5600
5601 CCGGTACCAC GCGAGTGT ACGACGGCA GGAGTTCACG GAGTACTGCC TCGCGAACT CGCCGCAAC ATCGACGAGA 5680
5681 TCGGCCCGG GCGGATCGCC CGATGATCG GGAGCGCGT CATGGCGCG GCGGCGCG TCGTCCCGC GCGGACTAC 5760
5761 TGGCCGCGG TCGCGCGCT GCTGCGCTCC CACGGCATCC TGCTGATCCT GGACGAGTC GTACCGCGT TCGGCGGCAC 5840
5841 GGGACCTGG TTGCGGCGG AGCACTTCG GTGACCCCC GATCTGTGG TGACCGGAA GGGCATCAC TCGGGTATG 5920
5921 TCCGACAGG GCGGTGCTC CTGACGAGG AGTTCGCGA CGCGTGAAC GGGGAGACG GTTCCCGAT CGGCTTCACC 6000
6001 TATACCGTTC ACCCACGCG GTGCGCGTC GCGCTCGCA ATCTCGACAT CATCGAACG GAAGGGCTGC TGGAGAACG 6080
6081 GTGAAGTGG GCGACACC TCGCCGGCG GCTGCGGCG CGCACCCCG CTGCGCGGC TGCCCGCGGT GGGGACGTC CGGCAACTGG 6160
6161 GCATGATGCT CCGGTGAG CTGTTGTGCG ACAAGAGGC CGCACCCCG CTGCGGGCG GCACCTCGG GGTGTTGGAC 6240
6241 GCGTGTGCG AGGACGCGG CGTCATGTC CGGCGCACG CTCGCGCGC TGGCACCCGA CGGCGGATC GCGCGGCC 6400
6321 GGCCACGCG GACGAGTGG CGGACGGGT GACTCGGTG GCGCGCGCG TCGGCACAGC GGCGACCG GGCCTTCCC 6480
6401 CCGGCGGG GTGACGAGC CGCGGCCGC CACCGCGGG ^{stop cvm6par} → start orf6par
6481 CGTTTCCCG CGCTTTTCC GTGCCCCGC GCGTTCCCG TGGCCCTGC CCTGCCCCT GTCGGGCGC TCCTCCCTCC 6560
6561 GTGTGCGC CGTTCCGTT CCAGCGGCT GTGAGCGCG CGCCAGGCG CCGTGCAC GGTGGAGAC CGCGGCCGA 6640
6641 CCGGGCGCG GAGGCCGCG AAGCCAGG TCCCGT CCGATCGCC TCCGACGTC CGCGCGGT CGGCGCGGT 6800
6721 ACGGTACGC CGGATCAG GGTTCACG CGGACCTCG GCGGACGC GGTCCCGAC GGGTCCCG GCCGCTGGT 6880
6801 TTCACCGGTT CGCGTTGCG CGCGCGAGT GTGTGCTCA GCGGAGCG ACGCCGCGA GTGCGCCAT CTGTGGCG 6960
6881 GTGTCTTCC GGCAACCCA ACGCGGAG GGGCCGCGG GGTACGAGG ACGCGGAGC GGTATCGAT GTCCGCTGC 7040
6961 GATCTGCA CTGACGAG AGGATGTC TGTGCGCTC CACGGACC GTGCGGAGC GGTATCGAT GTCCGCTGC 7120
7041 CCGGCCATC TCGGGCGGT GCGGGGCC TTACCGGTG CCGACTTCA CCGCGCGG GCGGCCGTG TGGCACCGC 7120

7121 GGGCGCCCGT CCCACGATCC GCGGGGCGG GTGCGGCGAC GCGACGCTGA TCGGTGTGCG CAAGGGCCCG GTACGGGCC 7200
 7201 CGGCGAGCA GAGACACCG TCGACGCTGG CGTTCTTCTG CACGAGCCCG CAGGTAGCC CGTCTGTCCT CGACGACATC 7280
 7281 TTCCGCCGGG TCGCGGACG CGCCTTCCAC GGGCTGGGCT TCGCGGCCGA CGCCTCCACC GCGCACACGG CGGCCGTTCT 7360
 7361 CGCCACGGG CTGCGGGCC GGGTGGACCT CGTCGGTTC GAACAGTTC TGGGCGCGCT GGGCTGGAC CTGCTCAGGG 7440
 7441 AGTCTGTCG GACACGGC TCGCGGGCG CCTTGGTAC GGTGCGGTC ACCGGGGCC ACGACACGA GCAGGCCGG 7520
 7521 CGGTGGGCC GGGCGGTGT CGACGGGCG TCGCTGAGG CCGCGGTGA CCGCCCGGCA CCGACTGGG CGCCGCTCGC 7600
 7601 CGCGTGGCG GGTGACACG GGGACGAAG CCGCGCCGG TCTCCCGGC GATCACGAT CCGGTGCGC GCGCGGAGG 7680
 7681 TCTTCCCGC CCCCAGCAC CGGGCCCGC CGACGCCGT TCCACGGCTG CAGCTCTCTG GCGGGGTACC CGGCTCTCG 7760
 7761 GACCTCGGTG TCCGGGCCG GCGGCCCGC GCGTTACCG GCGACCGCG GGGGCGGGA GCGCAGGGA CACGGGAGC 7840
 7841 CGCGGCCCG GCGTCTGAA ^{stop orf6par} GTCCCGGCC GCGTACCC CACCGCTGCC 8000
 7921 GCGCGGTG TCGATGGCC ACCGGGCCG CTCCGTCGT CGGTTCTCG GCGTTCTCG GTCGTCGCG CGCGGCCCC 8080
 8001 CGCGAGTC CAGGCGCTC TCGGCTCA CCGGCTCC TGGTGGCG GCGTACTGG TAGACAGTT CCGCCCGAT 8160
 8081 GGTGGCAGG GAGAGTCCAC CGGTCCGAC GCGGGCGAG GCGGGGACA CTTGACAC GTCGAGCCG ACGGGCCTGA 8240
 8161 CTCGCCCGC AGCAGGAGG TGAATCCCGA CGGTCGTAC GGACAGCCG CCGGGGCGG GTGTGCGGT GCCCGGGCG 8320
 8241 GCTGCCCGC CACGTCAGC AGGTCAGCA CTTGCGCGA GCGCAGGCC GCGCAGGTC GCGCGATCT GCTCGGCGAT 8400
 8321 TAGCCCGGT CGACGCTC GATGTGACG GAGACGTACA ATGCTGAGC GTGCGCGCG CGCGTAGTCC AGGAGTCGG 8480
 8401 GCGCGCGGT GAGCGCCGG TGAAGTCGC GCGGTGACG CTCGGGTCC ACCAGGCCCT CTTGATGGC CCAGCGGAG 8560
 8481 GCGCGGATT GTGGCGCG ATGCCGACT GAGACAGCG TGGTGTGCT GTGCGCTCC AGGTGACGA CCGCACCCG 8640
 8561 GGGGTGCGT GGTGTAGT GCCGCCGTAG ACGGTGGGT AGCAGTGT CCCCAGCCAG CATCAGAAC GCGTCTTTC 8720
 8641 GCGTGGCG GCGTCACCG CGGCGAGGC GCGCAGGAG CGCAGGTCC ATCGAAGG GCGTGGGTC GATGTGCCC 8800
 8721 GTTCCAGAG CCGGTGAG GCGACCTCG CGGTGTCCAT GGTGATGCC GACCGCTGG ATCAGGCTGG ACTGTGCG 8880
 8801 CGGTGACCA CTTGATCCG GTCGAAGACC CTTGGGCC GTGCTTCCGT CGTACGGGG GCGACGACC ACCAGTCAT 8960
 8881 GATGCGCG GCGCGAACC GCGCGCCGG CCGTAGCTG

8961 GGCCGATCGG GTCCGGCCGG TGGCGCAGCC GCATGAAGGT
 9041 CTGGCCGATC CC[GCGCACC CCGCCCTGCT CCGTTTCCG
 9121 GCCGACCCCG GTTCGCGAAC GGGCTCCCGT TCCCGCGTGG
 9201 TGCCTCTCC TCCGAGACCG CTCTGCTCGT TCCTGCGGCC
 9281 GCACGATCG CGCCGCGGCC GTTGCCGTTG CCGCGCGCCG
 9361 ACCACCGGC CCTGGAGCCT GAGCTTGGC ACCGCTCGA
 9441 CGGACCGTG TCGATGACCA CCGGTGCTGA CAGGCGCGT
 9521 TCCCTTCGGC GAGGAGCAGT CCGTTCACG CGCTGGTGT
 9601 GTCTCGGCG GCACGAGCC GGAGAGGCC TGCCTGACCA
 9681 GAGGAGACC AGCGGTCCA GGACCGCGG GTCCAGTAC
 9761 GGGACATCTC GTTAGGCGG TCGAGCCCG CCATGTGCC
 9841 CGGTGCATAC GCGGAGCGG GATGTGGGG CCTATACCGG
 9921 GCGGACGAGC GGCAGCAGT ACTCCAGGAC CGTGGGGTGG
 10001 CGAGTTCGGC CGAGTGGAGC CGGATCTCGC TGTGCGCGGT
 10081 TCGGACACCT CGGTGCCCAT CGACACGGAC CGTGTCCCGG
 10161 GATGCGCGCG GACAGGACGA CGGTGGGGC CGCTTCCCGG
 10241 CGACCAAGTC CACGCGCTCC CGTTGCGCGG GCAGCGCCCG
 10321 ATGTGCGAGC CGCGACTCC GTGCGAGCAGG AGGGCGGTCC
 10401 GTGGGTGCGG GACAGGCCCA GCGGCGGGC CGGCTCTGTC
 10481 CGTGGCGCGC CAGCCACAGC GGTACCGAAC CGGCTGTGTC
 10561 AGTGGCGGGA ACCGTCCGTT CAGGAGCCGG AAGGCCCGGG
 10641 GCGGAGGCGG GCAGAGGAGC CGCCGAGCGC TCCGTTACG

start orf4par ←
 CGCCGGTTGG GCGTAGCGG GGGAGACGGC GTGGACACC 9040
 TACCGACGC CGGCCAACC GTGGGGCTC CCGTTCCCGT 9120
 AATCCGCTTC CCGCGCCCG GCGCCGCTCC GGGCGCGCGC 9200
 GTTCCGCTC TCGGGCGCG TGCCTGCGC CAGCCCGCT 9280
 TCGCTCTG GCCACCGGTG CCGTTCTGGC CGTCTATACG 9360
 CCGAGCGCG CACGTTCTG CCGAAGTCCA CGTCTTCCGG 9440
 GGCATGCGC CCTGACGGC CGTCACCTCG TCGCGCGGA 9520
 GCGGACCCC TCGTGGATGC CCAGCTTGG GCGGGCCACG 9600
 CCCACTTGTG GTGCCCCGC CGCGTTTGA GCCCGGTTTC 9680
 GGTGGGTGG TCCACTTCCC GCGATGCCC GCAGGAGCG 9760
 CGCATCTCG TCGTCGAGG ACCAGACGA GGCCTGCGC 9840
 TCGGATGCG GAGCGGCCG GTCTCGAGC GCCGTAGAG 9920
 GTGATCTCG CCGCGCGGAC CGCCCAGGC AGTTCCCTGA 10000
 GCCCAGGTG ACGCGACCG AGCGGGCCG GTCGACTCG 10080
 GTGCCAGGC CGCGTGTGG GCGGCGACTC CCGCGAGTC 10160
 CCGCGAGCC GGTGCGGAC CGCGTGGCG AGCGTTGCG 10240
 GGAGAGCGG GGTGTCCAG TCGGACCGC CTGGCGGTG 10320
 CCGCGGGGAC CCGCAGAGC CCGCGCGCC CCGCGCGGT 10400
 GCCAGGCTCT TCGCTCGGT GCGGCGCTC AGCCCCGTCA 10480
 GTGGCGCGC ACCTGCGC CGTGGAGGC GTCGTGGAGC 10560
 GGGCCACGC CCGCAGGGG GCGCAGCA GTTCGGGCTC 10640
 TCGCGCGGT TGTACGCTC GCGCGCAGG AGCAGCCGGA 10720

10721 CCTGGCCGTC GCGACACCAG ACGGGCGGAC GGCCACGGGT CACGGCCGTT CCGCTCCAGA GCGGTACGC GGTGCCGTCG 10800
10801 TGCACGGGGA CATGGGTCCC GCGACAGGCG AAGCGGGTG CCGTCCGGG TTCCGAGTGA CCGCGGGGC CCGCGCGGG 10880
10881 GCGGCCCTCG GTGCCGATGC GCACCCGGAA TCCGTACACG AGTTCGGGGC CCGGCATGGT GAAATCGTCC TCGACGGTGG 10960
10961 TCGATGGCC AGGCGGCGA AACCGCCGGA CTGGAAGTCG TAGGCCACCG GTACCTCGAT CAGGACGGG CGGCCGAGTC 11040
11041 CCGGCCCTT GGTGAGGGG GCGAGCAGG AGGTGCGGTC GGTGGCGCGG ACGCCCTCGC AGCCGTTGGC CTCGGCGAGC 11120
11121 TGGACGAAGT CGACGCTTCC GAAGCCGACG GCGGGGCGT GCGAGGCGCTG GTGTCCGAGG TTCTGGTACA GCTCGATCAG 11200
11201 GCCGTTGCGG TCGTTGTGA CGACGACCAT GACGATCGGC AGGCCACAGC GCACGGCGGT CTCGATGTCG GCGCTGTTGG 11280
11281 AGTGAAGCC GCCGTGCCCC GCGATGAGGA AGACGGGCTC GCCGGGCGG GCGATCTGGG CCGCCATGGC GCGCGGCGAT 11360
11361 CCGTAGCCGA AGCTGGAGCA GCCCGCGGAG GTGAGGAATC CGTACGGCTG GTCGACTTG GCGAAGAGCA CGCCGTAGTG 11440
11441 GCGGAAGAG CCGATGTGCG TGACGAAGGT GCGGTTGTGCG AGGACGGAGT TCATGCATC GATCACCTGG TGGACCCGCA 11520
11521 TGCCGTCCTC GTACTCGGTG GGTGCGGCGA GGAATTCGGC GACGCGGGG CCGAGGGCGC TGAGGTGCTG CCGGCTCTTG 11600
11601 GGGGCGAGG CCGAGGTGCG GTGTCGAGC GCGGTGACGA ATTGCGCAC GTTGGTGAGC ATGTGATGT CGGCGCGGAA 11680
11681 CAGTCCGGG ATCGGTTGA CCTCGGGGC GACCCGACG GTGGTCTTGG CCGGCCCGG CBTCCACATG GAGGGGCGCA 11760
11761 GGTCTCGGC GTAGTCGTAG CCGATCGCCA GAGGAGGTC GCGGGGCGC AAGATCTGCT CGAGGCGCG GTGCCCCAGA 11840
11841 ATGCCGTCCA TGTAGCCGCT GATGGCCCG TACTTGAGCG GGTGGTCTG CCGCAGGACG CCTTTGGCGG TGTAGGTGGT 11920
11921 GACGACGGG ATGTTACGC GTCGCGGAG GCGCGCGAGG GCCTCGACG CCGCGCGCG GATGACGGG CTACCGAGCA 12000
12001 CGAGGAGGG GTTCTCGGCC TCGCGACCA GCTCAGCGGC CTCGTGAGG CCGCGCGCG AGTCGGCTC CAGGCGTGG 12080
12081 GTGGCGGTGG CCGGACACAG GGGGGCGTCG GTGGGGGTGC CGTTCAGTTC GGCGCGAGG AGTTCGACCG GCAGGCTGAT 12160
12161 GAAGCTGGGA CCCAGGGGT CGATCCGGCT GTTGAGGACG CGCTGTGCGA CGAGGTTGAC GATGCTCTCG CCGGTTTGA 12240
12241 GCTGGAGGCT GAATTTGGTC AGCGGGCCCA TCAGCGGGT GCTGTCCAG CACTGGTGGG TGAAGTTGGG GTAGCAGTCG 12320
12321 TACGACTCGG ACTGCGGCG CAGCGGATG ACCGAGCTGC GGTCCAGGGC GGAGTGGCG ACGCCGCTGG CCAGGTTGGT 12400
12401 CATGCCGGG CCCAGGTCG CGAAGCAGC CTGGGGGCGG TTGTTGATCC GGGCGAGGAC GTCCGCCATC ACCCGGCGG 12480
12481 TGAACTCGTG CCGGTCAGG ACGAAGTGA GTCCTTCGAC CTCGTCGAAG AAGTGGCGG ACGCCTCCG GCCGACGACG 12560

start of orf2par ←
12551 CCGAATACAT GGTGACACC GTACTGGTGA AGAGTTCCA GCATGGCTTT CGCGTCGTG GTGGCCATGG AGATCTCCTT 12640
12641 CGCATCGGAC GGGCGCCGGG ATGGCCCCC GGAACAACGG GCACCGGCG GTGCGACCG GTGGGCGCAC ACCGTGGGTG 12720
12721 GTGGCGTTGC CACTGTGCG ATGCCCTCTT GCGGCGGTC GGACGCCCG CTTGACAGA ATGGGCAAGG CGCGTTCAAG 12800
12801 GCATGGGTC CATCGTCCTC GTGGCGCTTT TCGTGAATC CGTCGGGCG CGACGCTC CATCCGATC GTCCCCCTTC 12880
12881 CGTCCACCGA TCCGAGGAGA ATCCATGGAT GTCTGGCGG CGTTGGAGCG CAAGCCAGC CTGAATCTTT TCCCATCGA 12960
12961 GAACCGGTG TCGCCGCGG CCAGTGCCG GCTGGCCACC GACGCCGTC ACCGCTATCC GTACTCGAG ACCCGGTGG 13040
13041 CGCTCTACGG CGATGTCAG GGGCTGGCG AGGTGTACGC GTACTGCCAG GACCTGGCCA AGCGCTTCTT CGGGGCGGCG 13120
13121 CACGCCGGTG TGCAGTTCTT GTCCGGTCTG CACACCATGC ACACGCTGCT GACCGCCCTG ACCCGGCCG GCGGGCGGCT 13200
13201 CCTGGTCTC GCGCCGGAGG ACGGCGGCCA CTACGCCAGG GTGACGATCT GCCGGGGCTT CGGCTACGAG GTCGAGTTCT 13280
13281 TACCTTCGAC CGCCGACAC CTGGAGATCG ACT 13313
| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80

PGT/EP004/004001



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.